Bio-Physical Models of Oceanic Population Dynamics

1994 Summer Study Program

19980416 113

DTIC QUALITY INSPECTED 4

**Course Lectures
Abstracts of Participants
Fellows Project Reports**

WHOI-97-18

Bio-Physical Models of Oceanic Population Dynamics
1994 Summer Study Program in Geophysical Fluid Dynamics

by

Glenn Flierl and Donald Olson, co-Directors

Woods Hole Oceanographic Institution
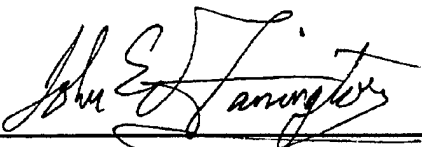Woods Hole, Massachusetts 02543

November 1997

Technical Report

Approved for Distribution:

John W. Farrington
Associate Director for Education
and Dean of Graduate Studies

Contents

Contents (continued)

1994 GFD Fellows, Staff and Visitors

The Fellows

Martin Bees	Leeds University, UK
Emmanuel Boss	University of Washington
Tonya Clayton	University of South Florida
Amanda Cohen	University of Washington
Amar Gandhi	Princeton University
Maryam Golnaraghi	Harvard University
Deborah Le Bel	University of Washington
Douglas Morgan	Thayer School of Engineering, Dartmouth College
Michael Neubert	University of Washington
Scott Stewart	University of Colorado

Staff and Visitors

James Anderson	Stevens Institute of Technology
Robert Armstrong	Princeton University
Neil Balmforth	University of Texas, Austin
Pavel Berloff	Florida State University
Robert Cantrell	University of Miami
Hal Caswell	Woods Hole Oceanographic Institution
Eric Chassignet	University of Miami
George Cosner	University of Miami
Glenn Flierl	Massachusetts Institute of Technology
Mick Follows	Massachusetts Institute of Technology
Robert Frazel	Woods Hole Oceanographic Institution
Daniel Grunbaum	University of British Columbia, CANADA
John Guckenheimer	Cornell University
Louis Howard	Florida State University
George Hurtt	Princeton University
Joseph Keller	Stanford University
Sergey Kraustov	Florida State University
Simon Levin	Princeton University
Willem Malkus	Massachusetts Institute of Technology
Arthur Mariano	University of Miami
Richard Matear	Institute of Ocean Sciences, CANADA
Salvatore Mazzola	Massachusetts Institute of Technology
Dennis McGillicuddy	Woods Hole Oceanographic Institution
Stephen Meacham	Florida State University
Donald Olson	University of Miami
Mercedes Pascual	Woods Hole Oceanographic Institution
Nathan Platt	Naval Warfare Research Center

Staff and Visitors (continued)

Thomas Powell	University of California, Berkeley
Michael Renardy	Virginia Polytechnic Institute and State University
Yuriko Y. Renardy	Virginia Polytechnic Institute and State University
Kathy Rhodes	University of California
Claes Rooth	University of Miami
Andrew Solow	Woods Hole Oceanographic Institution
Edward Spiegel	Columbia University
John Steele	Woods Hole Oceanographic Institution
Melvin Stern	Florida State University
Shripad Tuljapurkar	Morrison Institute for Population and Resources
Matthew Umurhan	Columbia University
George Veronis	Yale University
Jack Whitehead	Woods Hole Oceanographic Institution
Andrew Woods	University of Cambridge, UK
Hidekatsu Yamazaki	Tokyo University of Fisheries, JAPAN
Philip Yecko	Columbia University
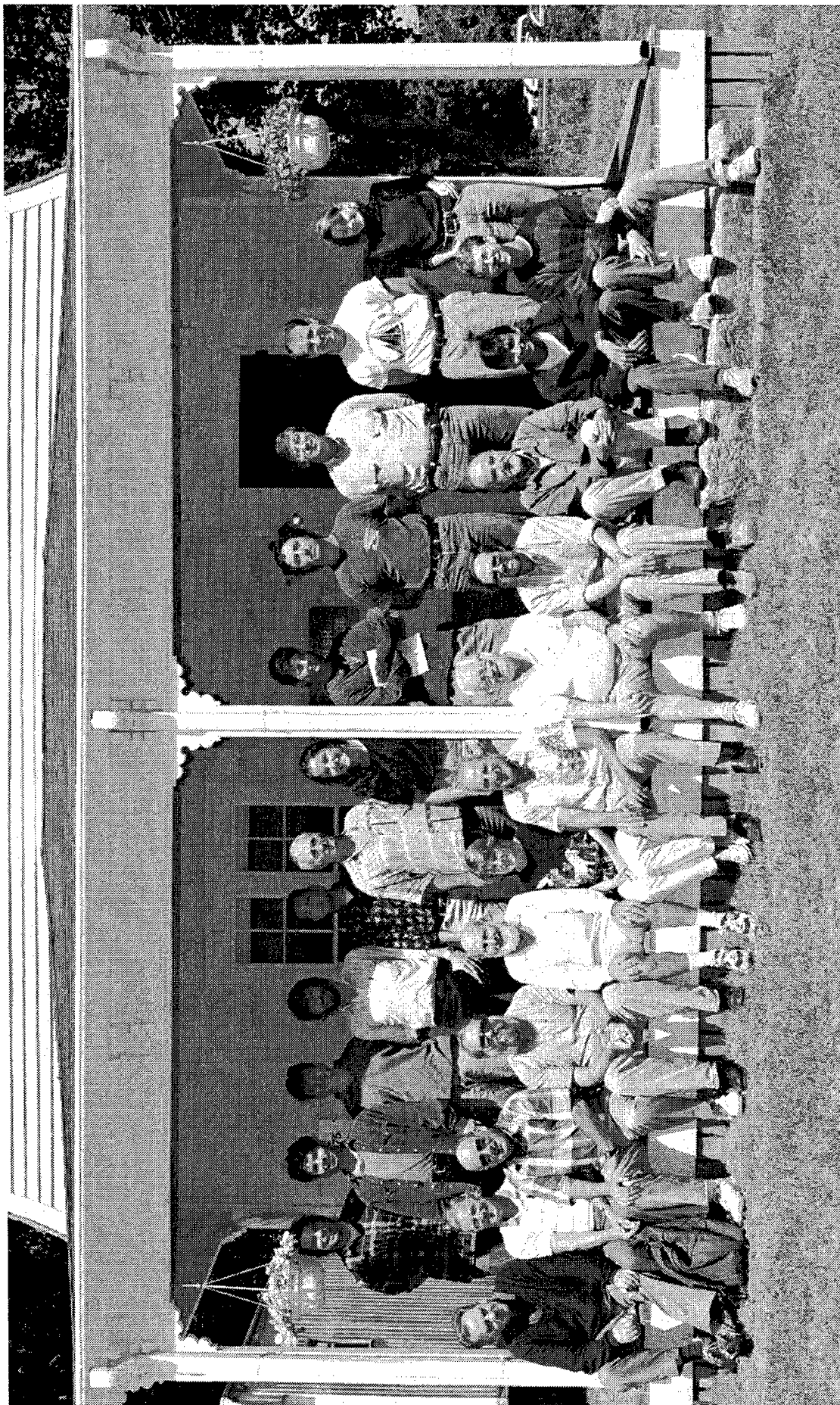
1994

Standing (l-r): M. Umurhan, M. Bees, N. Balmforth, S. Meacham, P. Yecko, W. Malkus, A. Cohen, D. Grunbaum, E. Boss, S. Mazzola, M. Neubert, D. Le Bel
Seated (l-r): E. Spiegel, J. Whitehead, R. Frazel, D. Olson, J. Keller, M. Thurston, G. Veronis, C. Rooth, G. Flierl, M. Stern, A. Gandhi, S. Stewart
Missing from photo: J. Anderson, R. Armstrong, P. Berloff, R. Cantrell, H. Caswell, E. Chassignet, T. Clayton, G. Cosner, M. Follows, M. Golnaraghi, J. Guckenheimer, L. Howard, G. Hurtt, S. Kraustov, S. Levin, A. Mariano, R. Matear, D. McGillicuddy, D. Morgan, M. Pascual, N. Platt, T. Powell, M. Renardy, Y. Renardy, K. Rhodes, A. Solow, J. Steele, S. Tuljapurkar, A. Woods, H. Yamazaki

vi

GFD 1994 Lecture Schedule

Introduction to Oceanic Biology and Physics
Monday, June 26, 10 AM
D. Olson

Age Structure and Complexity in Biological Models
Tuesday, June 21, 10 AM
G. Flierl

Basic Formulations of Animal Aggregation and Movement
Wednesday, June 22, 10 AM
D. Grunbaum

Dispersal of Populations: Terrestrial vs. Fluid Environments
Thursday, June 23, 10 AM
D. Grunbaum

Models for Predator--Prey Systems at Multiple Scales
Thursday, June 23, 2 PM
S. Cantrell

Individual Behavior and Population Patterns
Friday, June 24, 10 AM
D. Grunbaum

Data for Demographic Plankton Models
Friday, June 24, 2 PM
J. Steele

Evolutionary and Adaptive Trade-offs
Monday, June 27, 10 AM
D. Grunbaum

SEFCAR
Monday, June 27, 2 PM
C. Rooth

TBA
Tuesday, June 28, 10 AM
D. Grunbaum

Tutorial on GFD
Tuesday, June 28, 1:30 PM

Matrix Population Models
Wednesday, June 29, 10 AM
H. Caswell

Flows and Biology
Thursday, June 30, 10 AM
G. Flierl

Tutorial on Oceanic Biology
Thursday, June 30, 2 PM

Parameter Optimization and Analysis of Ecosystem Models
Friday, July 1, 10 AM
R. Matear

Bioconvection
Tuesday, July 5, 10 AM
E. Spiegel

Computing Bifurcations in Dynamical Systems
Wednesday, July 6, 10 AM
J. Guckenheimer

TBA
Thursday, July 7, 10 AM
T. Powell

Populations in Heterogeneous Environments
Friday, July 8, 10 AM
S. Levin

Scale as an Ecological Problem
Monday, July 11, 10 AM
S. Levin

Chemical Waves
Tuesday, July 12, 10 AM
L. Howard

Tutorial on Mixed Layer Dynamics
Tuesday, July 12, 1:30 PM

Modelling Biogeochemical Interactions in the Subtropical Gyre
Wednesday, July 13, 10 AM
M. Follows

Mesoscale Ocean Dynamics and Biological Productivity
Thursday, July 14, 10 AM
D. McGillicuddy

On-Off Intermittency
Friday, July 15, 10 AM
N. Platt

Explosive Decompression and Eruption of Magmas
Friday, July 15, 2 PM
A. Woods

Principal Component Analysis of Plankton Patterns
Monday, July 18, 10 AM
A. Mariano

Finite Amplitude 3-D Instability of Inviscid Laminar Boundary Layers
Tuesday, July 19, 10 AM
M. Stern

Structure of Oceanic Turbulence in Channel EGCMs
Thursday, July 21, 10 AM
D. Olson

Biological Structure and Spatial Dynamics; or, the Homogeneous Green Stuff Fights Back
Friday, July 22, 10 AM
S. Tuljapurkar

Survival -- What is it and When Does it Occur?
Tuesday, July 26, 10 AM
G. Cosner

The Statistical Analysis of Population Time Series Data
Wednesday, July 27, 10 AM
A. Solow

Influence of Outcropping Layer on Mid-Latitude Jet Separation
Thursday, July 28, 10 AM
E. Chassignet

Flow of Water and Carbon Through Bryozoan Colonies
Monday, August 1, 10 AM
D. Grunbaum

Modelling the Trophic--Dynamic Aspect of Ecology
Tuesday, August 2, 10 AM
R. Armstrong

From Cell Cycles to Population Cycles in Phytoplankton Dynamics
Thursday, August 4, 10 AM
M. Pascual

Aspects of Spatio-Temporal Complexity
Friday, August 5, 10 AM
N. Balmforth

Interfacial Instabilities in Viscous Stratified Flows
Thursday, August 11, 10 AM
Y. Renardy

An Eigenvalue Problem Arising in the Study of the Stability of Ocean Currents
Friday, August 12, 10 AM
M. Renardy

Turbulence and Plankton
Monday, August 15, 10 AM
H. Yamazaki

Some Problems in Bio-Mathematics
Friday, August 19, 10 AM
J. Keller

FELLOWS' LECTURES

Instability and Pattern Formation in a Discrete--Time Predator--Prey Model
Monday, August 22, 3 PM
M. Neubert

Effects of Nutrients Storage and Light Adaptation in Phytoplankton Ecology
Tuesday, August 23, 9:30 AM
E. Boss

Cline Formation in a Population Genetics Model
Tuesday, August 23, 11 AM
A. Cohen

A First Look at NPZD Models
Wednesday, August 24, 9:30 AM
S. Stewart

The Birds and Bees
Wednesday, August 24, 11 AM
M. Bees

Pattern Formation in Slime Molds
Wednesday, August 24, 3 PM
A. Gandhi

Propagation of Isolated Eddies on a Beta-Plane with Sloping Topography
Thursday, August 25, 10 AM
D. Le Bel

Toward an Improved Dispersion Operator for Organisms with Long Mean Free Paths
Thursday, August 25, 1:30 PM
D. Morgan

Marine Population Dynamics: A theoretical introduction

Donald B. Olson

The purpose of this lecture is to lay out the basic problems in population dynamics in a concise manner and compare them with the equations for the fluid dynamics that they must be coupled with to understand marine ecosystems. To begin let us define a population of organisms that have characteristics allowing them to be pooled into a single population based on genetics, *i.e.* species or lineages within a species, or functional considerations such as all herbivorous zooplankton in a region. This population can then be quantified as the ith population in some spatial/temporal domain within a larger scale ecosystem in terms of its population density, n_i. This measure of population maybe expressed in the number of individuals per volume or in some "biogeochemical currency" such as μmols - nitrogen per unit mass in the environment as a whole. The latter is common in a wide variety of models that one can express in terms of nitrogen pools, for example the inorganic nitrogen, phytoplankton, zooplankton or NPZ model (Olson and Hood , 1994; Steele and Henderson, 1991). The dynamics of such a population are expressed in terms of all of the other relevant ecosystem components by an equation of the form

$$dn_i/dt \;=\; F_i(\mu_{ij}, n_j), \quad i = 1, 2, 3, ...m.$$

Here μ_{ij} represents the typically spatial variable parameters that govern a population's reactions with itself and the other m components of the ecosystem. F_i is an operator that describes the interaction of the parameters and the m populations. Prior to exploring the nature of F_i it is worthwhile to introduce an alternate form,

$$dn_i/dt \;=\; n_i \, F_i(\mu_i j, n_j)$$

where the population density is explicitly expected to appear at least linearly in all of the terms on the right. This equation is known as the Kolmogorov equation after the illustrious Russian contributor to both fluid and population dynamics.

The details of a population's behavior described by these equations depends on the specification of F_i. In particular in the case that $F_i \neq F_i(t)$ the populations are said to consist of a autonomous system involving an endogenous set of coupled populations, *i.e.* the system is a closed dynamics system. Under this situation one might consider the population as a whole and ask for equilibria

corresponding to the solutions of $F_i(\mu, n_j) = 0$. These equilibrium states or fixed points, $\hat{n}_i$, can be further analyzed through methods similar to those used in fluid dynamics to consider the stability of populations. The fluid dynamics analog here are the Euler equations for a flow

$$d\vec{u}/dt = -\frac{1}{\rho}\nabla p.$$

Systems with temporal dependence in the operator form non-autonomous and therefore non-Hamiltonian systems governed by equations of the form

$$dn_i \ dt = F_i(\mu_i, n_j, v(\gamma, t))$$

where $v(\gamma, t)$ represents a forcing or dissipative term. Such systems are referred to as exogenous systems. The fluid analog is the Navier-Stokes equation

$$d\vec{u}/dt = -\frac{1}{\rho}\nabla p + \nu\nabla^2\vec{u}.$$

The details of population dynamics and the manner in which it couples with the oceanic circulation depends on the terms implicit in the total derivatives (d/dt) and the specification of F_i. First of all it is assumed that the derivative is a material one following the flow. Therefore the derivative on the left includes advective effects that can be taken to include both the mean flow, $\vec{U}$, and turbulent diffusion. Writing the derivative out gives

$$dn_i/dt = \vec{U} \cdot \nabla n_i + \nabla\overline{\vec{u}'n_i'}$$

The last term is typically included as a diffusion term whose correct specification will be a topic in many of the following lectures. On the right hand side in the F_i term the self-motive forces of populations can be written in similar terms. In this form one can take the motion on the part of the biology and add it to the material derivatives. That is, the total population motion will be governed by the vector addition of the flow and their swimming, $\vec{u}_{total} = \vec{u} + \vec{u}_{swimming}$. Similarly the diffusion of a population will involve both behavior and the flow. The biological diffusion will in general depend on the other populations in $F_i(n_j)$ in the sense that predator-prey or density dependent interactions within a population such as competition for resources or reproduction will add to diffusion. Since these interactions are at a minimum spatially dependent "Fickian" diffusion terms will typically appear in the form $\nabla \cdot (K_i\nabla n_i)$ at least. The problem of specification of the diffusivity, K_i, in terms of animal behavior will be treated in later lectures by Danny Grübaum.

Beyond the effects imparted on populations by their own ability to move the remaining interactions between organisms and their environment involves the interplay between reproduction, *i.e.* population growth, and mortality or population loss. These can be written explicitly in an expansion of our operator

$$F_i(t, \mu, n_j) = G_i(\mu_g, n_j) + M_i(\gamma, n_j)$$

where the population growth operator, G_i, is typically a nonlinear functional that includes both linear growth terms and what are known as density dependent terms that can be written in the form $\beta_{ij} n_i n_j$. The β_{ij} is a competition coefficient governing the interactions between populations. Population growth in the fisheries literature is often treated as a recruitment term that accounts for reproduction in a manner that often includes the factors following actual birth that limit the arrival of new members into the population. This is an example where the population is defined by a species at a given age when individuals are recruited to the population of concern to the fishery. Often fisheries models are written in terms of power curves to reflect the population density influences on population growth.

Mortality in populations can either assume a proportional rate of death or may have some form of predator-prey relating consumption of one population by another in the ecosystem. Let us first examine the simplest form of both growth and mortality,

$$dn/dt = gn - dn = g_I n$$

where $g_I = g - d$ is the net population growth. The difficulty with this equation is evident from its fixed point, $\hat{n} = 0$, and its solution, $n = n_o e^{g_I t}$. The latter shows that the equilibrium point is stable with no population for $g_I < 0$ and becomes infinite for $g_I > 0$. This is the Malthusian dilemma (Malthus, 1798) that was explored by a number of scientists including Graunt, Linnaeus, Franklin, and Euler in the 17th and 18th century (see Caswell, 1989). The conclusion is that this simple linear equation can not realistically depict a population.

Instead to achieve realistic populations the governing equations must contain higher order terms. This leads to the formulation of what is known as logistic growth (Verhulst, 1836)

$$dn/dt = rn(1 - n/K)$$

where r is the maximum growth and K the carrying capacity of the environment. This equation has fixed points $\hat{n} = K, 0$ and the time dependent solution

3

$$n = \frac{n_o K e^{rt}}{K + n_o(e^{rt} - 1)}.$$

Therefore the solution for any finite initial n asymptotes to K *i.e.* the $\hat{n} = 0$ solution is unstable. Several cases with simple logistic dynamics are shown in Fig. 1.

The logistic equation as written above is a competition model and as such does not really account for mortality. A more complete parameterization, then, might take the form

$$dn/dt = rn(1 - n/K) - \frac{d\,n^2}{(k^2 + n^2)}$$

that has been used to simulate spruce bud worm population dynamics by Ludwig *et al.* (1978). The death term here is parameterized as a sigmoid curve (Fig. 2). The student may consider the following questions concerning this system.

- How many equilibrium solutions exist for this equation?

- What is the non-dimensional form and structure of the solution?

For a full treatment of these questions see chapter one of Murray (1990). There are a number of nonlinear forms that have been suggested in the literature as closures for population models. These are summarized in Fig. 2. Mathematically each parameterization produces its own twist on the dynamics of a population. The first, Lolkta-Volterra is in essence a rediscovery of the 18th century debate on population closures mentioned above. The Michaelis-Menton form derives from studies of bacteria cultures in the middle of this century. The Ivlev formulation is a favorite in fisheries biology, but in the author's experience gives rather strange fixed states. The sigmoid curve of the spruce bud worm problem is perhaps the best formulation for both predator-prey and uptake problems. Unfortunately, it is also a problem in terms of solving for fixed points since it generally leads to polynomials higher than five, even in very simple models such as the NPZ. Asymptotic parameterizations that include the behavior of the sigmoid function as a dn^2 term at low population levels ($k^2 >> n^2$) have been strongly suggested by Steele and Henderson (1991).

It is worth discussing at this point the degree to which we can discern the appropriate choices based on observations. This is the single largest challenge in biological oceanography today. While huge steps are being made to better quantify various biomass quantities and to determine rates, much of the present work is incapable of deciding between formulations. In particular, there is a major debate on whether or not there is any sort of density dependent control on marine populations. The discussion

4
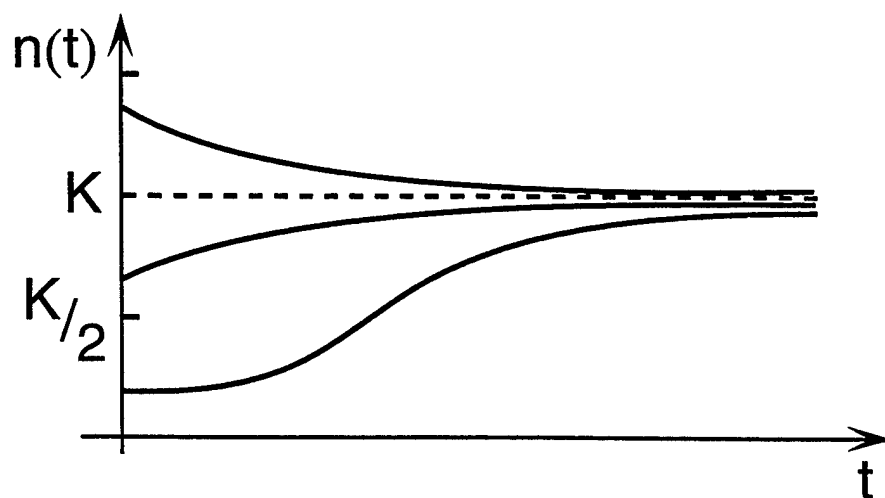
Fig. 1. Solutions to the logistic equation for population density, $n(t)$, with initial conditions above the carrying capacity, K, of the environment, just below K and below $K/2$. Note that all of the cases asymptote to K in time, t.
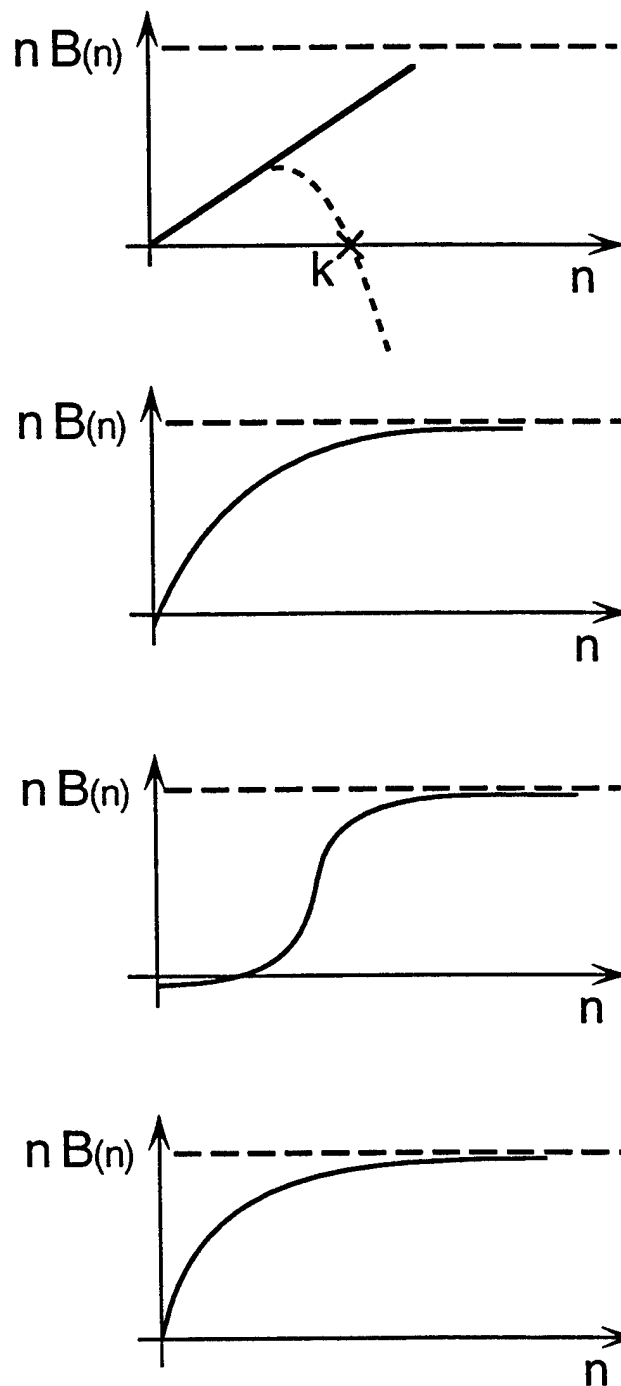
Fig. 2. Population dynamic parameterizations commonly used in the literature. From top to bottom these are the linear formulation with and without a carrying capacity term (dashed). The second is the exponential function introduced by Ivlev. The third is the sigmoid or Hollings type III curve. The fourth is the Michaelis-Menton formulation that can be derived from chemical dynamics and is often used in relationship to nutrient uptake in phytoplankton. This is also referred to as a Hollings type II.

6

above points out that there must be some form of density dependent closure on populations. Yet, with the current observational capability there is little evidence that it acts in real populations. This either suggests that populations are maintained in a range where this sort of behavior does not act, in our parlance here they exist in a linear range of the equations, or the observations are inadequate to discern the effects of density dependence. This leads us now to a set of linearized population models.

Linearized Pooled Models

There are various places one might start to model marine ecosystems. A common choice pursued in various coastal programs such as the Coastal Upwelling Ecosystems Analysis (CUEA, Wroblewski, 1977) and by investigators interested in carbon or nutrient fluxes such as those working on Joint Global Ocean Flux Studies (JGOFS) modeling is a system that is written in terms of some "currency" variable which is conserved within the system. The nitrogen, phytoplankton, zooplankton (NPZ) model discussed by Steele (1974), written in terms of nitrogen content as the biogeochemical "currency", is the basic example of this approach. An extended version of this sort of system is the one developed by Fasham *et al.* (1990). Here let us consider a simple linearized version of NPZD or nitrogen, phytoplankton, zooplankton, detritus model. The basic set of equation are:

$$\dot{N} = a(1-\gamma)gPZ + (1-\delta)dZ + eD - uIPN$$

$$\dot{P} = uIPN - gPZ - sP$$

$$\dot{Z} = a\gamma gPZ - dZ$$

$$\dot{D} = (1-a)gPZ + \delta dZ + sP - eD.$$

The variables are defined below. The entire system along with an expansion to explain the partition of nitrogen during zooplankton grazing are shown in Fig. 3.

$N \equiv$ Nitrogen concentration in μmoles/kg.

$P \equiv$ Phytoplankton concentration in nitrogen equivalent (μmole s /kg).

$Z \equiv$ Zooplankton concentration (μmoles/kg).

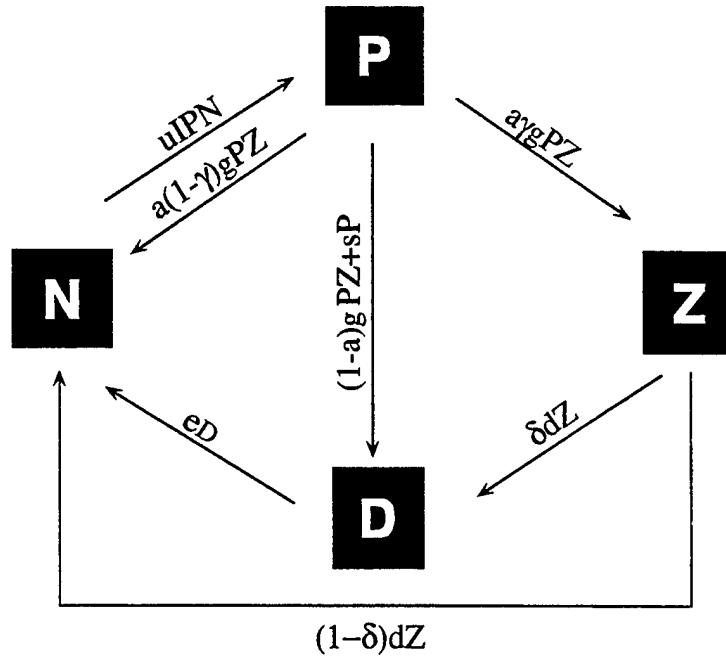$D \equiv$ Detritus concentration (μmoles/kg).

Fig. 3. Structure of the NPZD model for plankton dynamics. See text for details.

$I \equiv$ Light intensity. This is a function of depth and in general P.

$u \equiv$ Phytoplankton growth rate per unit N and I.

$g \equiv$ Grazing rate.

$d \equiv$ Death rate for zooplankton.

$e \equiv$ Remineralization rate for detritus.

$s \equiv$ Senescence term for phytoplankton.

$a \equiv$ Assimilation efficiency for Z. Ratio of assimilation to ingestion.

$\gamma \equiv$ Metabolic efficiency for Z. Ratio of growth to assimilation.

$\delta \equiv$ Assimilation efficiency for higher trophic levels.

Here the dot ($\dot{\,}$) is meant to denote the time derivative. Since we want to treat populations in relation to ocean flows this derivative will in general be replaced by the material derivative ($d/dt = \partial/\partial t + \vec{V}\cdot\nabla$) and a diffusion term will be added to the right-hand side of the equations. In an initialization of the model it is necessary to specify the total amount of nitrogen in the system or

$$N_o = N + P + Z + D.$$

The problem can be generalized through the use of indicial notation and expressed in terms of a matrix. In this form the equations above can be written as

$$\dot{B}_i = \mathbf{L}_{ij}B_j + \mathbf{N}_{ik}B_iB_k + f_i$$

where B_i denotes the ith component and $\mathbf{L}_{ij}$ and $\mathbf{N}_{ik}$ are linear and nonlinear operators representing the terms on the right above. The system above can expressed in matrix form.

$$
\begin{bmatrix} \dot{P} \\ \dot{Z} \\ \dot{N} \\ \dot{D} \end{bmatrix}
=
\begin{bmatrix}
-s & 0 & 0 & 0 \\
0 & -d & 0 & 0 \\
0 & (1-\delta)d & 0 & e \\
s & \delta d & 0 & -e
\end{bmatrix}
\begin{bmatrix} P \\ Z \\ N \\ D \end{bmatrix}
$$

$$+ P \begin{bmatrix} 0 & -g & uI & 0 \\ 0 & a\gamma g & 0 & 0 \\ 0 & a(1-\gamma)g & -uI & 0 \\ 0 & (1-a)g & 0 & 0 \end{bmatrix} \begin{bmatrix} P \\ Z \\ N \\ D \end{bmatrix}$$

Note that all of the nonlinearity involves the P-pool with either N or Z. There are no nonlinear terms in D which might arise if an active bacteria pool was explicitly included. As a general comment, however, the detritus pool is isolated in the models such as Fasham *et al.* (1990) and can easily be collapsed into the current form. The largest effect of adding bacteria is to make the *e* time dependent. Finally, while the system is nonlinear it does not include logistic nonlinearities. For example it might be necessary to include explicit density dependent interactions between pools. The most common means of expressing these involve terms such as

$$\hat{u}IP(N - P)/(K_o + N) ; \quad a\gamma\hat{g}Z(P - Z)/P,$$

for the phytoplankton and zooplankton growth terms respectively where $\hat{u}$ and $\hat{g}$ are the new growth coefficients. These are the Michaelis-Menten (M-M below) and Lotka-Voltera (L-V below) forms. Note that these terms are not easily expressed in our generalized form above and will involve prognostic variables, N, Z, P in the matrix.

Equilibrium Solutions

The NPZD equations can have up to four real, positive equilibrium solutions. The relevant solutions ($B_i \geq 0$.) can be found by setting all of the time derivatives on the left of the equations equal to zero and then solving for the equilibrium conditions, $\hat{B}_i$. The first equilibrium state has

$$\hat{P} = d/a\gamma g$$

from the Z equation. Then $\hat{Z} = (uI\hat{N} - s)/g$ from the P equation. Using the total nitrogen to specify $D = N_o - \hat{P} - \hat{N} - \hat{Z}$ and substitution gives

$$\hat{N} = (as(1 - \gamma)d/a\gamma g + (1 - \delta)ds/\gamma + e(d/a\gamma g - s/g - N_o))/A$$

where

$$A = uIa(1 - \gamma)d/a\gamma g + (1 - \delta)duI/g - uId/a\gamma g - e(1 + uI/g).$$

10

Back substitutions then give $\hat{D}$ and $\hat{Z}$. The second equilibrium occurs as $\hat{N} \to s/uI$ and therefore from the solution above as $\hat{Z} \to 0$. In this case $\hat{D} = \hat{P}s/e$ and

$$\hat{P} = (N_o - s/uI)/(1 + s/e).$$

The final equilibrium of interest occurs as $N_o \to s/uI$ and $\hat{P} = \hat{D} \to 0$. The three solutions are shown graphically in Fig. 4 as functions of N_o and uI. These are equivalent to the three states identified by Flierl and Davis (1992) for the simpler system without D. Interestingly the addition of D adds no equilibrium where the other variables are greater than zero.

The lack of $\hat{P}$ dependence on light in the initial equilibrium has bothered various authors (*e.g.* Steel and Henderson, 1992). In fact the dependence on light only enters the P terms in the transition between the various equilibria contrary to what intuition suggests. In the area above the $\hat{Z}$ extinction curve in Fig. 4. $\hat{P}$ is independent of light and grazing controlled. It then becomes a function of light in the second region. The comments of Steele and Henderson (1992) concerning this highly linearized set of equations are, however, well taken and it is worthwhile to explore these before attempting to generalize the pooled approach to multi-species pools. The analysis above is therefore primarily useful as an illustration of how a more complete system might be formulated. The equilibrium states in models are crucial for understanding the system.

Multi-species Formulations

The equations above can be expanded to include any number of species. Here they are expanded to include two phytoplankton and two zooplankton. The equations become

$$\dot{P}_1 = u_1 I P_1 N - g_{11} P_1 Z_1 - s_1 P_1$$

$$\dot{P}_2 = u_2 I P_2 N - g_{12} P_2 Z_1 - g_{22} P_2 Z_2 - s_2 P_2$$

$$\dot{Z}_1 = a_{11} g_{11} P_1 Z_1 + a_{12} g_{12} P_2 Z_1 - d_1 Z_1 - g_z Z_1 Z_2$$

$$\dot{Z}_2 = a_{22} g_{22} P_2 Z_2 + a_z g_z Z_1 Z_2 - d_2 Z_2$$

$$\dot{N} = -u_1 I P_1 N - u_2 I P_2 N + \Delta_{11} P_1 Z_1 + \Delta_{12} P_2 Z_1 + \Delta_{22} P_2 Z_2$$
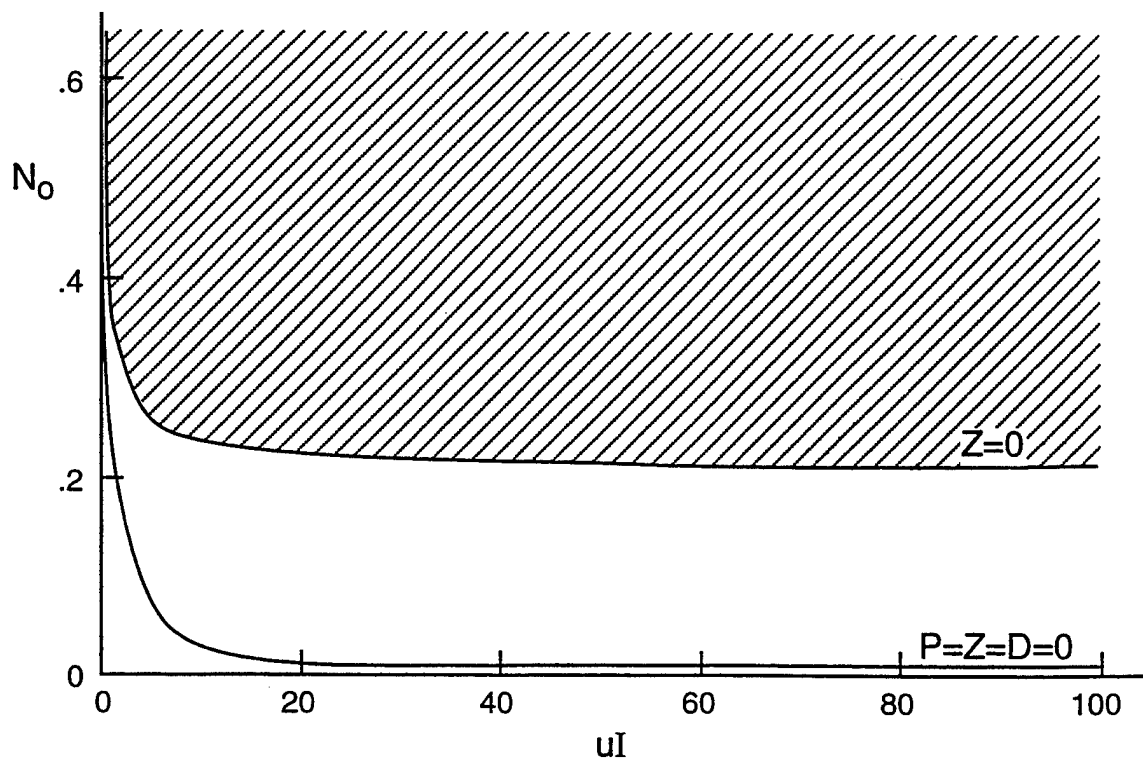
11

Fig. 4. Regime diagram for the fixed states of the NPZD model. The hatched area indicates the region
in total nutrient and light space (N_o vs uI) where all four variable have finite concentrations. At
the curve labeled $Z =$) the zooplankton become extinct. At the lower curve P goes to extinction
carrying D with it. Below this curve the nutrient is all in the inorganic pool.

$$+ \Delta_z Z_1 Z_2 + (1 - \delta_1)d_1 Z_1 + (1 - \delta_2)d_2 Z_2 + eD$$

and

$$\dot{D} = \alpha_{11} P_1 Z_1 + \alpha_{12} P_2 Z_1 + \alpha_{22} P_2 Z_2$$

$$+ \alpha_z Z_1 Z_2 + \delta_1 d_1 Z_1 + \delta_2 d_2 Z_2 + s_1 P_1 + s_2 P_2 - eD$$

where $\Delta_{ij} = (a_{eij} - a_{ij})g_{ij}$, $\alpha_{ij} = (1 - a_{eij})g_{ij}$ and the c_z terms denote coefficients for zooplankton feeding on zooplankton. All of the terms have the same definitions as above. Note that this formulation can be used to describe two parallel food chains one envisioned to consist of pico-phytoplankton (P_1) and micro-zooplankton (Z_1) and the other containing larger phytoplankton (P_2) and something like copepods (Z_2). Here a set of terms allowing grazing of the smaller phytoplankton by the Z_2 component has been dropped. In its place the chains are linked by allowing grazing of the Z_1 by the larger zooplankton.

As above the equilibrium states, $\hat{B}$ can be found by setting $\dot{B}_i = 0$. Writing down just the $\hat{P}$ and $\hat{Z}$ solutions in terms of $\hat{N}$ gives

$$\hat{Z}_1 = \{u_1 I \hat{N} - s_1\}/g_{11}$$

$$\hat{Z}_2 = \{u_2 I \hat{N} - g_{12}\hat{Z}_1 - s_2\}/g_{22}$$

$$\hat{P}_2 = \{d_2 - a_z g_z \hat{Z}_1\}/(a_{22}g_{22})$$

$$\hat{P}_1 = \{d_1 + g_z \hat{Z}_2 - a_{12}g_{12}\hat{P}_2\}/(a_{11}g_{11}).$$

Solution of the N and D equations with constant total nitrogen (N_o) in the system yields a quadratic equation for $\hat{N}$. Note that these equilibria have explicit light dependence in the phytoplankton solutions. Therefore adding more realistic food-web dynamics leads to more intuitive light dependence without the Z^2 closure of Steele and Henderson (1992). Finally, this system has four extinction states that are all of the form $\hat{B} \to 0$ as $I\hat{N}$ decreases to a constant made up of the parameters in a form similar to those in the simple NPZD case. The extinction horizons are then hyperbolic curves in

13

the $\hat{N} - I$ space. This is analogous to the simple NPZD system where the extinction horizons are hyperbolic in $N_o - I$.

There is not time today to go into the influence of non-linear dynamics on food chains. Students are encouraged to consider the model of Hastings and Powell (1991) who consider a single food chain with density dependent terms and show that it exhibits limit cycles and chaos.

References

Caswell, H. 1989. *Matrix Population Models: Construction, analysis and interpretation.*, Sinauer Assoc., 328 pp.

Fasham, M. J. R., H. W. Ducklow, and S. M. McKelvie 1990. A nitrogen-based model of plankton dynamics in the oceanic mixed layer. J. Mar. Res. **48**, 591-639.

Hastings, A. and T. Powell 1991. Chaos in a three-species food chain. *Ecology*, **72**(3), 896-903.

Ludwig, D., D. G. Aronson, H. F. Weinberger 1979. Spatial patterning of the spruce budworm. *J. Math. Biol.*, **8**, 217-258.

Malthus, T. R. 1798. An essay on the principal of population.

McCann, K and P. Yodzis 1994. Biological conditions for chaos in a three-species food chain. *Ecology*, **75**, 561-564.

Murray, J. D., *Mathematical Biology*, Springer-Verlag, New York, 767 pp.

Olson, D.B. and R.R. Hood, Modeling Pelagic Biogeography. Prog. Oceanogr., 34, 1994, pp. 161-205.

Steele, J. H. and E. W. Henderson 1992. The role of predation in plankton models. *J. Plankton Res.*, **14**, 157-172.

Verhulst, P. F. 1838. Notice sur la loi que la population suit dans son accroissement. *Corr. Math. et Phys.*, bf 10, 113-121.

Wroblewski, J.S., 1977. A model of phytoplankton plume formation during variable Oregon upwelling. *J. Mar. Res.*, **35**, 357-393.

Lecture 2: Complexity in Biological Models
Glenn Flierl, June 21, 1994

Von Foerster's Equation

Biological systems are complex in many ways. Ecosystems are comprised of many species, with different and changing life strategies. Organisms can interact as predator-prey, or symbiotically, or as competitors. Even individuals in a single species behave differently as their age and as their internal states (e.g., nutritional) change. Individuals and propagules move spatially and, in the ocean especially, are subject to movements from physical flows. To illustrate some of the complexities and dynamics found in biological systems, we shall consider a single species and discuss the distribution of weights in the population and how it changes with time. This is called the Von Foerster–McKendrick equation.

Derivation

Let $n = n(w, x, t, n)$ be the number of individuals of a certain species per unit weight, where w denotes weight, and x and t denote space and time. Then the total number of individuals is given by:

$$\int_{w_0}^{w_1} n \, dw,$$

where w_0 and w_1 are the minimum and maximum weight of individuals. The number of individuals between w and $w + dw$ changes with time from gains and losses as individuals increase in weight, moving into or out of the weight class, and by death:

$$n(w, t + dt)dw = n(w, t)dw - g(w + dw)n(w + dw, t)dt + g(w)n(w, t)dt - \mu(w)n(w, t)dt$$

15

where g (which can be expressed as dw/dt) is the growth rate, and μ is the death rate. In the limit as dt and dw go to zero, this equation can be expressed as follows:

$$\frac{\partial n}{\partial t} = -\frac{\partial}{\partial w}(gn) - \mu n \tag{1}$$

(NOTE: Equation (1) is analogous to the equation for conservation of a tracer in a flow, with g replacing $\bar{u}$, and μn acting as a sink. The g is inside the derivative just as tracer per unit volume would be in a conservation equation. We usually see terms like $u\frac{\partial}{\partial x}S$ because we deal with tracer per unit mass while u is the volume flux rather than the mass flux.)

The boundary condition for equation (1) is the production of individuals in the lowest weight class. Assuming constant weight at birth, this production is given by:

$$g(w_0)n(w_0) = \int_{w_0}^{w_1} b(w',t)n(w',t)dw' \tag{2}$$

where $b(w',t)$ is the weight specific birth rate (the reproductive function).

I) Linear Solution:

Assuming that g and μ are not functions of n, we can solve equations (1) and (2) as follows: Rewriting in terms of gn, we find that:

$$\frac{\partial(gn)}{\partial t} = -g\frac{\partial}{\partial w}(gn) - \mu(gn) + \frac{1}{g}\frac{\partial g}{\partial t}(gn)$$

and using the method of characteristics,

$$\frac{dw}{dt} = g(w,t)$$

$$g(w(t),t)\ n(w(t),t) = g_0 n_0 \exp\left[-\int_0^t \left(\mu - \frac{1}{g}\frac{\partial g}{\partial t}\right)_{w(t'),t'} dt'\right]$$

where $g_0 = g(w(0),0)$ and $n_0 = n(w(0),0)$ are the initial growth rate and population distribution, respectively. Holding the birth rate fixed, we can evaluate the steady state solution:

$$\frac{\partial(gn)}{\partial w} = -\frac{\mu}{g}(gn)$$

16

or

$$gn = g_0 n_0 \exp\left(-\int_{w_0}^{w} \frac{\mu}{g} dw'\right)$$

This solution is illustrated in Fig. 2, for the functions g and μ shown in Fig. 1. As shown in Fig. 2, n is a minimum where g is a maximum, and vice versa. This result arises because the time an individual spends in a given weight class is inversely proportional to the growth rate of that weight class.

II) "Turbulent" Solution:

Allowing g and μ to fluctuate randomly around given means shifts the mean distribution of n slightly (Fig. 3). To explain this effect, we can decompose n, g and μ into mean (ensemble average) and fluctuating quantities:

$$n = \bar{n} + n', \quad g = \bar{g} + g', \quad \mu = \bar{\mu} + \mu',$$

and substitute these expressions into (1) to derive equations governing the time evolution of $\bar{n}$:

$$\frac{\partial \bar{n}}{\partial t} = -\frac{\partial}{\partial w}\overline{gn} - \overline{\mu n} - \frac{\partial}{\partial w}\overline{(g'n')} - \overline{\mu' n'} \tag{4}$$

and that of n':

$$\frac{\partial n'}{\partial t} = -\frac{\partial}{\partial w}\left(\bar{g}n' + g'\bar{n} + g'n' - \overline{g'n'}\right) - \bar{\mu}n' - \mu'\bar{n} - \mu'n' + \overline{\mu'n'} \tag{5}$$

These equations are analogous to those governing turbulent advection of a passive tracer. As in the fluid dynamics analysis, there is a closure problem for equations (4) and (5). To get some idea of the effects of the fluctuations, we assume that $\bar{g}$ and $\bar{\mu}$ are constants and that g' and μ' are functions of time only. We then linearize equation (5):

$$\frac{\partial n'}{\partial t} + \bar{g}\frac{\partial n'}{\partial w} + \bar{\mu}n' = -g'\frac{\partial \bar{n}}{\partial w} - \mu'\bar{n}$$

We can find each of the particular solutions separately (after some algebra) and conclude that

$$\overline{\mu'n'} = \int_0^\infty Cov_{\mu'\mu'}(\tau)\bar{n}(w - \bar{g}\tau, t - \tau)e^{-\bar{\mu}\tau} d\tau \longrightarrow \overline{\mu'^2}\tau_\mu \bar{n}(w, t)$$
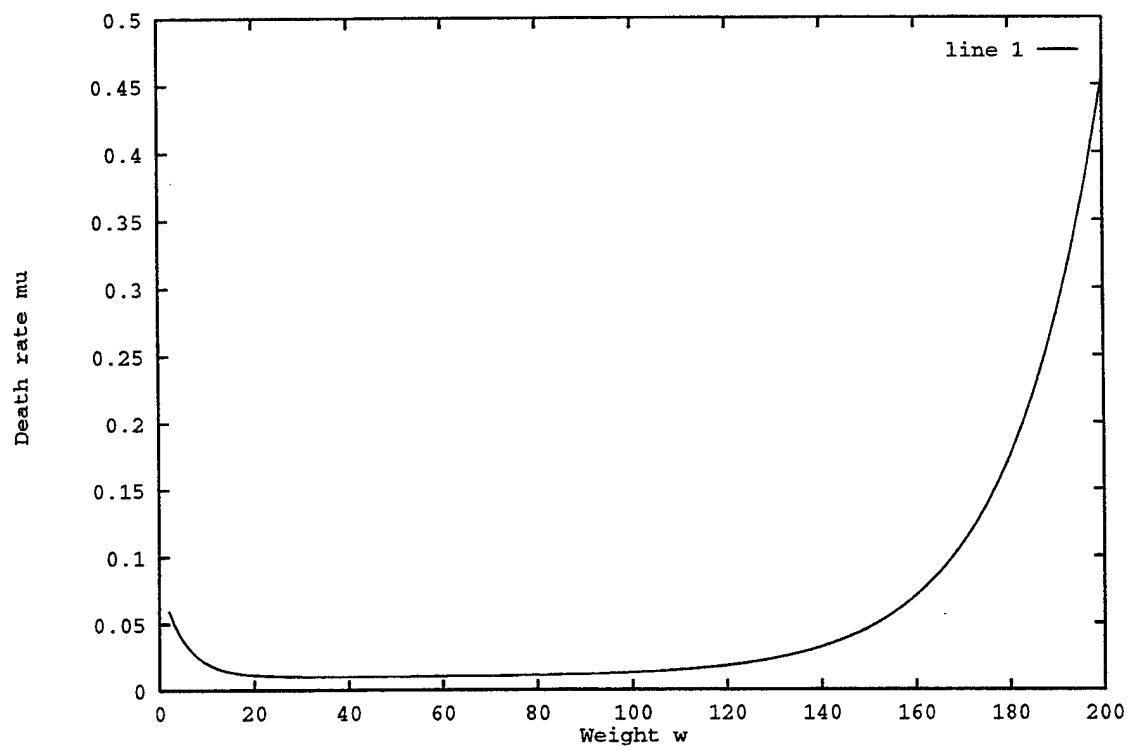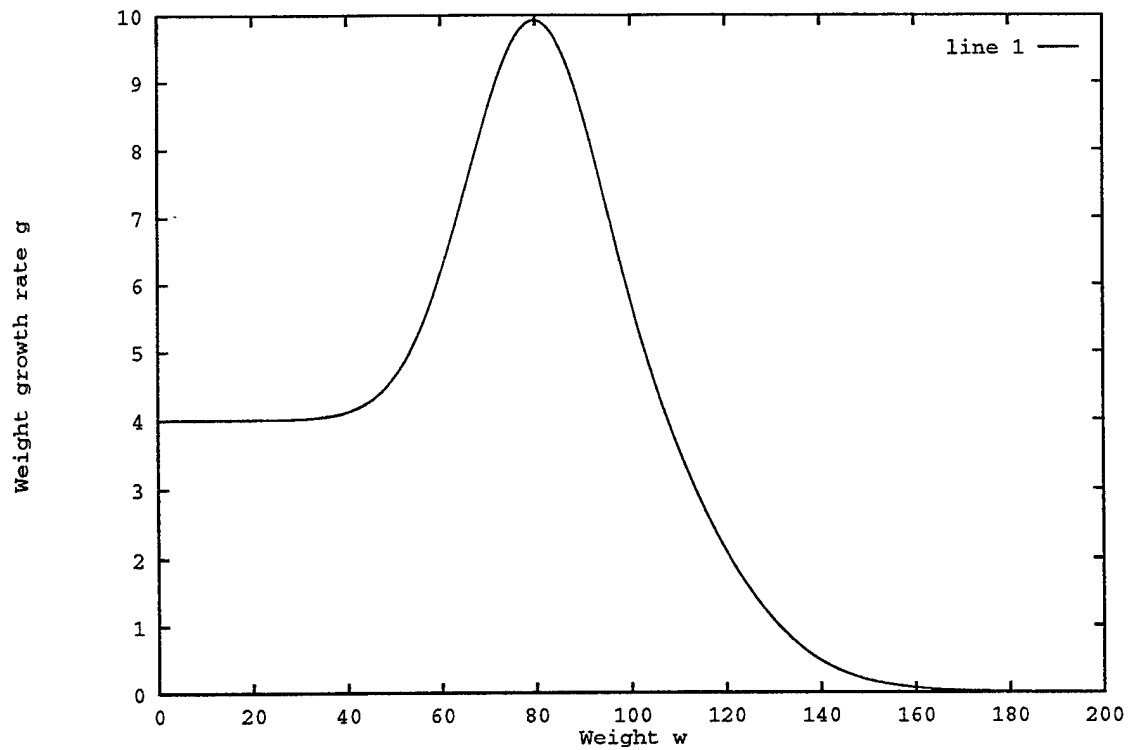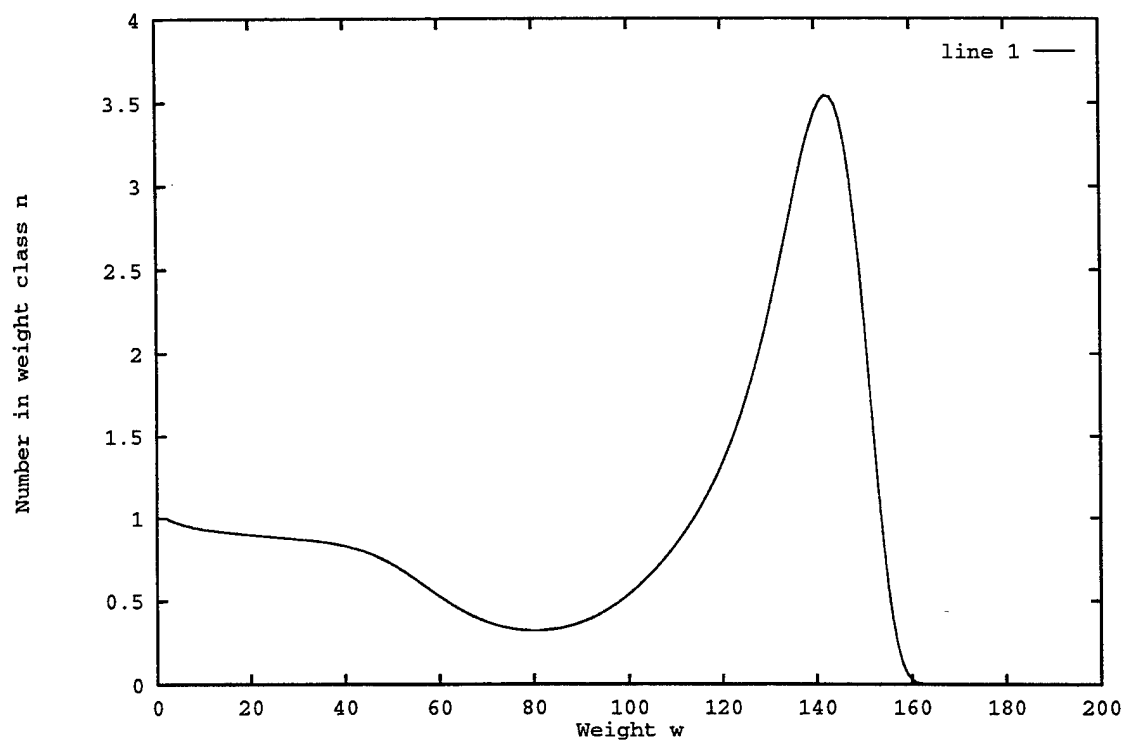
17

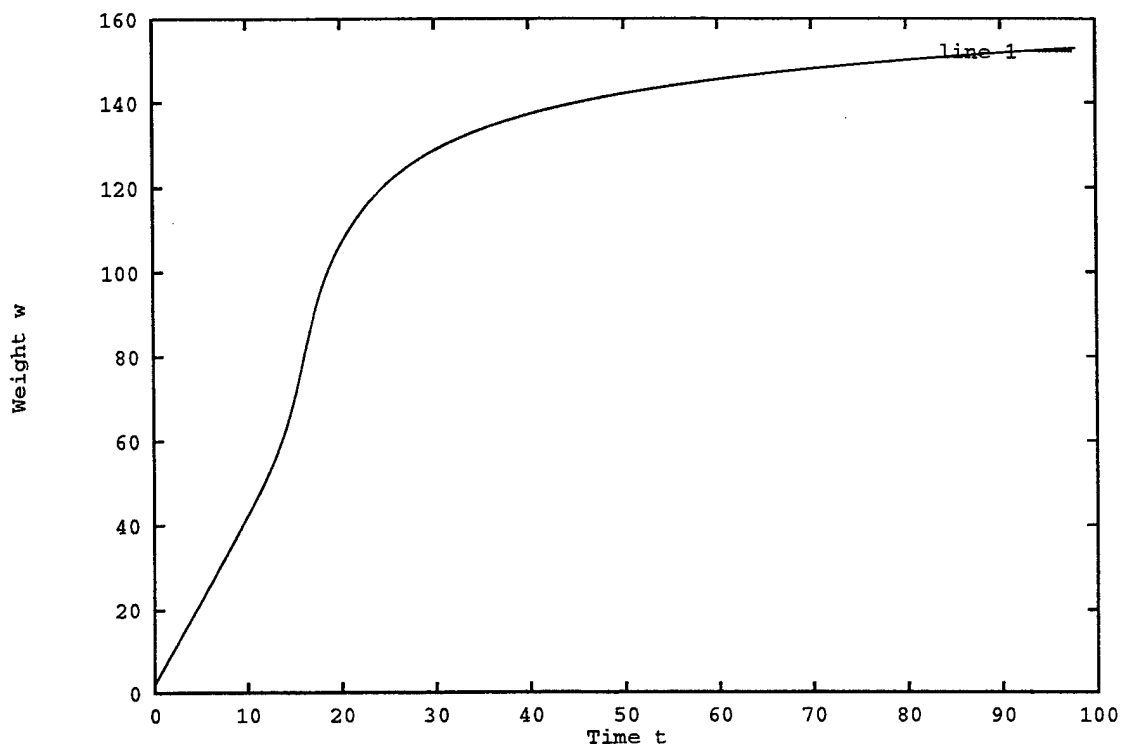Figure 1: Weight growth and death rates
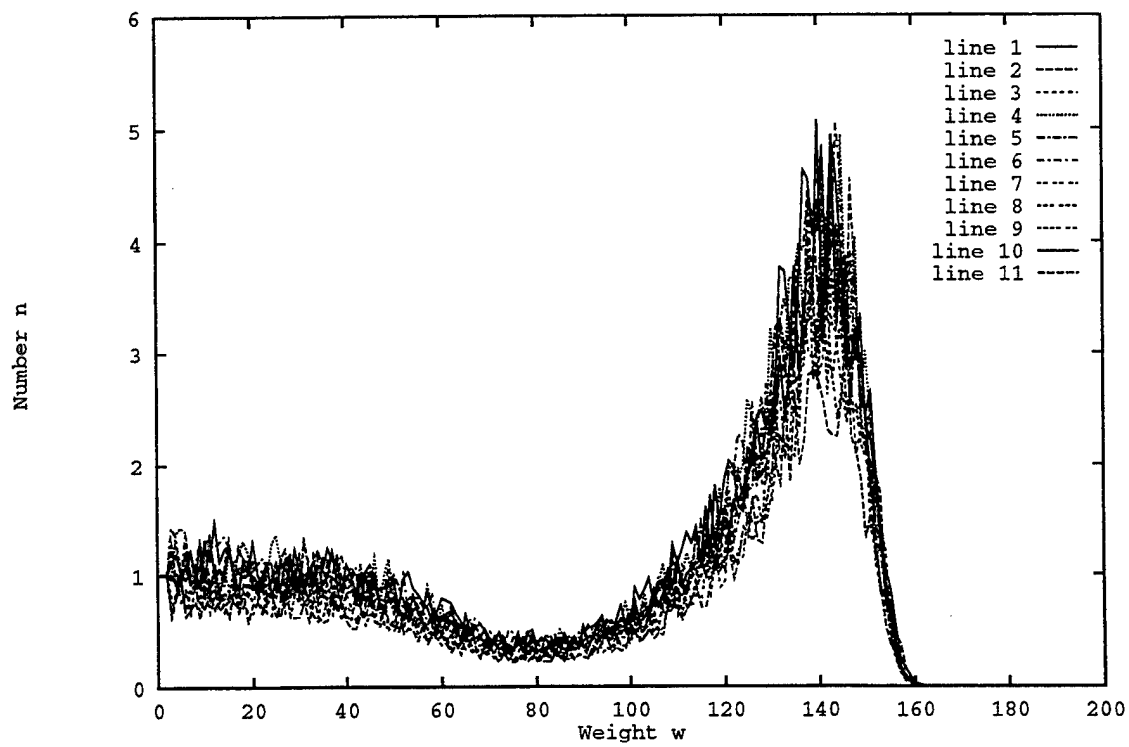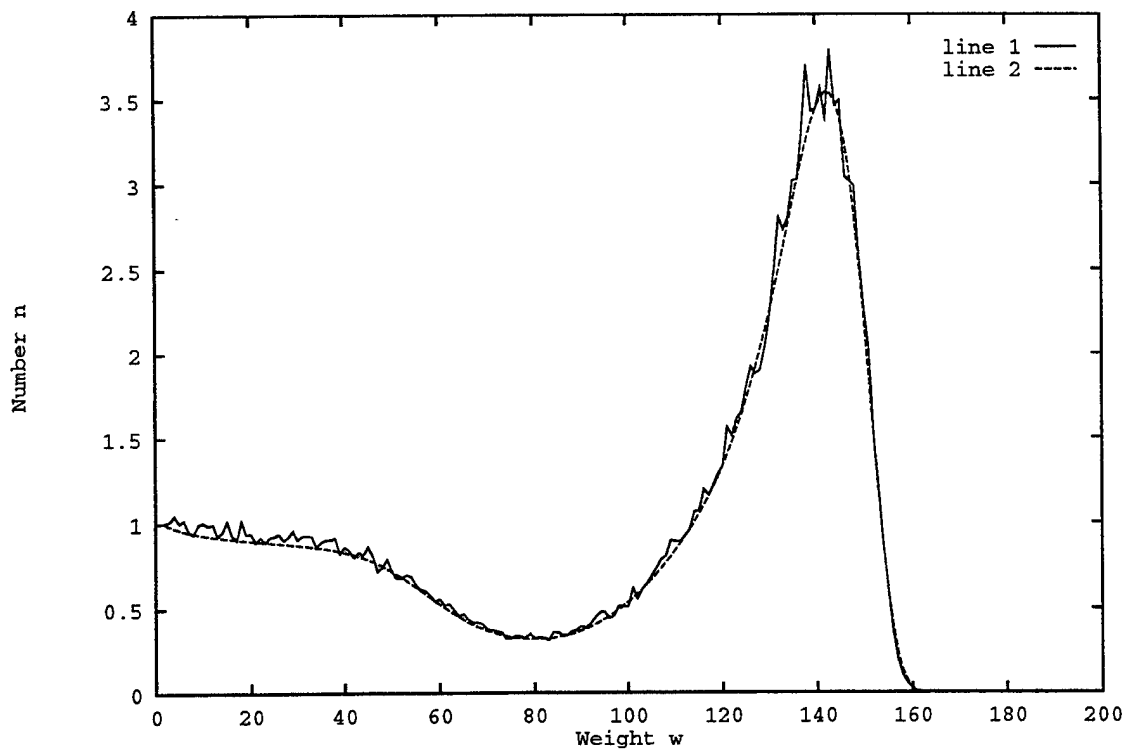
18

Figure 2: Weight vs. time and number vs. weight

Figure 3: Number vs. weight in 10 realizations and mean, compared to steady case

$$\overline{g'n'} = \int_0^\infty \text{Cov}_{g'g'}(\tau)\frac{\partial \bar{n}}{\partial w}e^{-\bar{\mu}\tau}d\tau \longrightarrow -\overline{g'^2}\tau_g\frac{\partial \bar{n}}{\partial w}(w,t)$$

where the decorrelation times τ_μ and τ_g are assumed short. The terms $\left(\overline{\mu'^2}\tau_\mu \bar{n}(w,t)\right)$ and $\left(-\overline{g'^2}\tau_g\frac{\partial \bar{n}}{\partial w}(w,t)\right)$ can be substituted back into equation (4) to determine a final expression for the time rate of change of the mean population distribution, $\bar{n}$. The first term, $\left(\overline{\mu'^2}\tau_\mu \bar{n}(w,t)\right)$, increases the death rate, μ, (a parameter change) while the second term, $\left(-\overline{g'^2}\tau_g\frac{\partial \bar{n}}{\partial w}(w,t)\right)$, gives a flux proportional to the gradient (in weight) — it acts to diffuse the mean distribution of individuals (a structural change).

III) Nonlinear Dynamics

The dynamics (1–2) will be nonlinear if g, μ, or b depend in some way upon the population numbers, whether total or in particular weight classes. In order to understand the nonlinear behavior we first calculate the linear modes of equations (1) and (2). We assume $g = 1$ so that weight (w) is equivalent to age (a). Then equation (1) becomes:

$$\frac{\partial n}{\partial t} = -\frac{\partial n}{\partial a} - \mu n \tag{6}$$

with the boundary condition that

$$n(0,t) = \int_0^\infty b(a)n(a,t)da \tag{7}$$

This system has the solution (for μ constant):

$$n(a,t) = n(0,t-a)e^{-\mu a} \tag{8a}$$

where

$$n(0,t) = \int_0^\infty \left[b(a)n(0,t-a)e^{-\mu a}\right]da \tag{8b}$$

We then look for a normal mode structure for $n(0,t)$:

$$n(0,t) \propto e^{\sigma t}$$

21

Plugging this form into (8b) gives

$$1 = \int_0^\infty b(a)e^{-(\mu+\sigma)a}\,da \tag{9}$$

The values of σ satisfying equation (9) depend on the form of the reproductive function, $b(a)$.

Examples:

i. If individuals only reproduce at a specific age, a_0:

$$b(a) = b_0\delta(a - a_0)$$

Plugging this reproductive function into (9) we find

$$1 = b_0 e^{-(\mu+\sigma)a_0} \Rightarrow \sigma_m = -\mu + \frac{1}{a_0}\left(\ln(b_0) + i2\pi m\right) \quad, \quad m - integer$$

(Figure 4) where the multiple σ_m's result from the different branches of the natural log in the complex plane. Notice that all the complex growth rates (σ_m) have the same real part.

ii. For a more general reproductive function centered at a_0 we can rewrite (9) as:

$$1 = b_0 e^{(\mu+\sigma)a_0} \int_0^\infty b(a)e^{-(\mu+\sigma)(a-a_0)}\,da$$

Taylor expanding the integrand:

$$e^{-(\mu+\sigma)(a-a_0)} \approx 1 - (\mu+\sigma)(a-a_0) + \frac{1}{2}(\mu+\sigma)^2(a-a_0)^2\ldots$$

and assuming that $b(a)$ is symmetric around a_0:

$$1 \approx e^{(\mu+\sigma)a_0}\left[\int_0^\infty b(a)\,da + \frac{1}{2}(\mu+\sigma)^2\int_0^\infty (a-a_0)^2 b(a)\,da\right] \tag{10}$$

$$\equiv e^{(\mu+\sigma)a_0}\left[B + \frac{1}{2}(\mu+\sigma)^2 D\right]$$
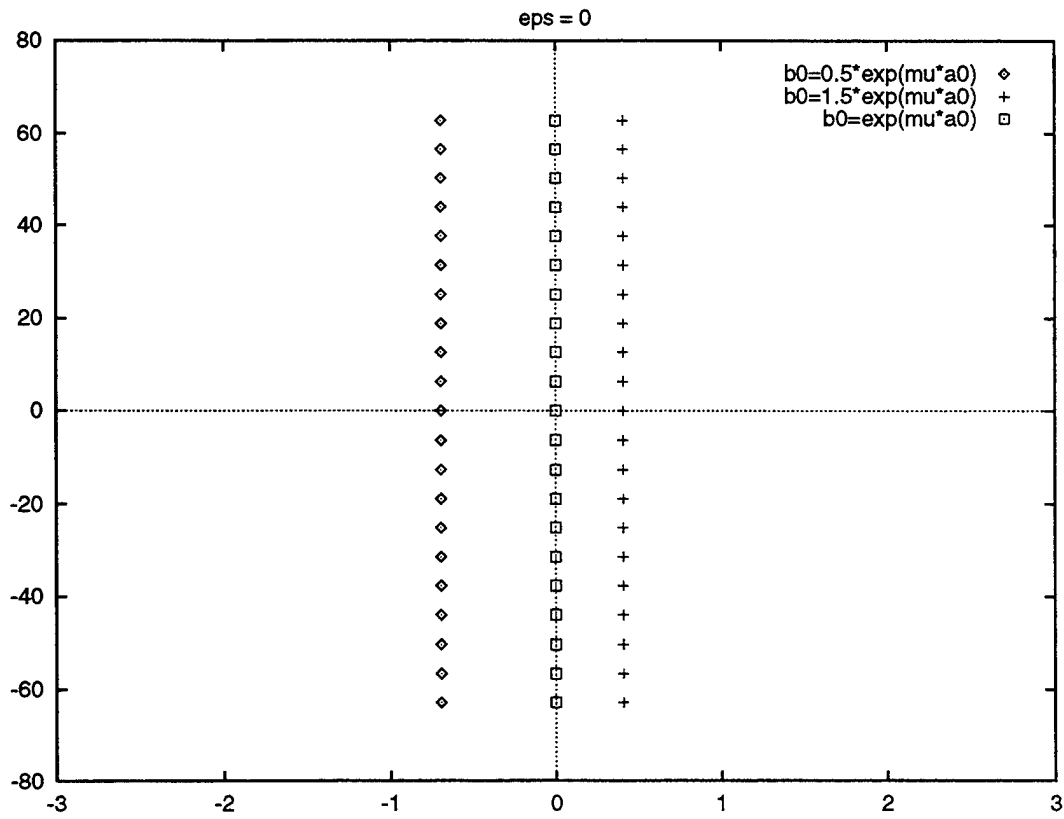
22

Figure 4: Growth rates $\sigma_m a_0$ for various b_0
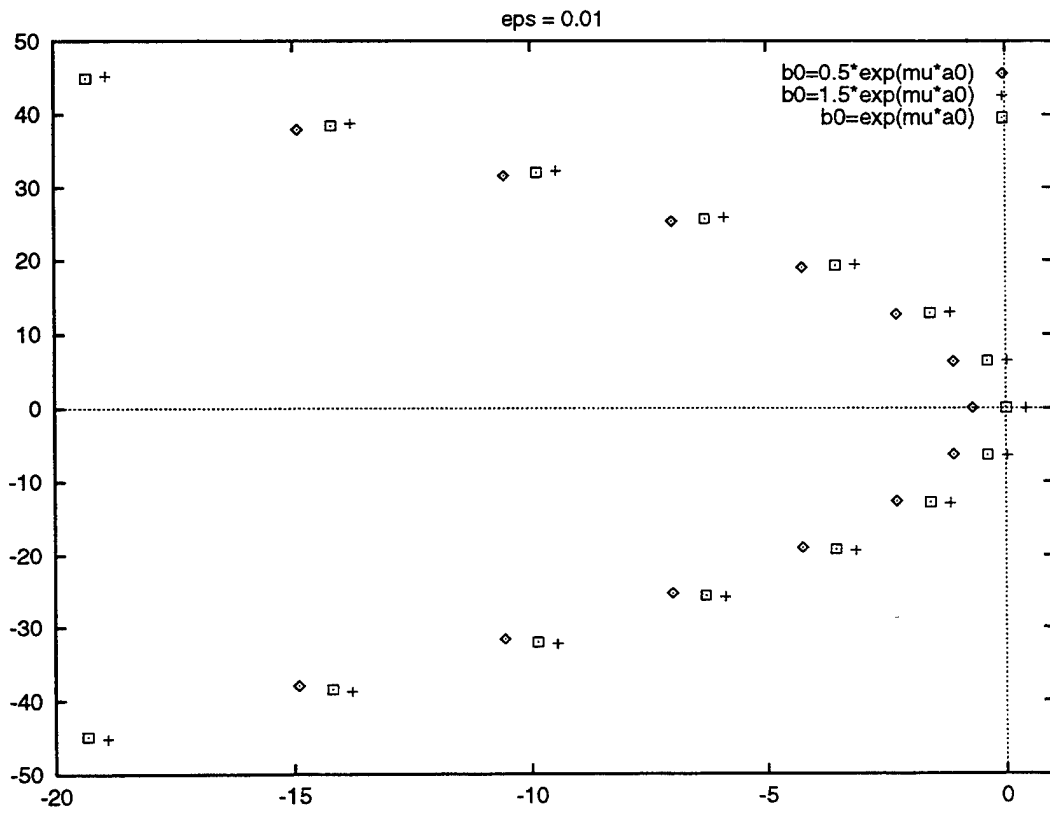
Figure 5: Finite width reproduction curve, with $\epsilon = 0.01$

23

For a reproductive function narrowly distributed around a_0, we recognize that $D << B$. Defining ϵ as the ratio (D/B), we can expand σ:

$$\sigma = \sigma_0 + \epsilon\sigma_1 + ...$$

and Taylor expand to find

$$e^{-(\mu+\sigma)a_0} \approx e^{-(\mu+\sigma_0)a_0}[1 - \epsilon\sigma_1 a_0 + ...]$$

The $O(1)$ balance in equation (10) is

$$1 = Be^{-(\mu+\sigma_0)a_0}$$

which has the same solution as in the delta function case considered above. The $O(\epsilon)$ balance is then

$$\sigma_1 a_0 = (\mu + \sigma_0)^2$$

Solving for the growth rate, $Re(\sigma)$, gives

$$Re(\sigma_m) = -\mu + \frac{\ln(B)}{a_0} + \frac{\epsilon}{a_0}\left[\left(\frac{\ln(B)}{a_0}\right)^2 - \left(\frac{2m\pi}{a_0}\right)^2\right] \quad, \quad m - integer$$

From this equation, we can see that the growth rate decreases as $|m|$ increases (Figure 5). For $m = 0$, σ is real and the growth rate is a maximum. Over time, this real σ dominates the solution for $n(0, t)$, and the population approaches a fixed age distribution.

The general theorem:

For the linear problem, if $b(a)$ has a finite width, there is a stable age distribution toward which the system tends.

Leslie Matrix

Returning to equation (6)

$$\frac{\partial n}{\partial t} = -\frac{\partial n}{\partial a} - \mu n$$

24

we can discretize the system to give

$$n(a,t) = n(a - \Delta a, t - \Delta t) exp \left(- \int_{a-\Delta a}^{a} \mu da' \right) ,$$

or

$$\bar{n}_{t+\Delta t} = \begin{bmatrix} 0 & & & 0 \\ v_1 & 0 & & \\ & \ddots & \ddots & \\ 0 & & v_{m-1} & 0 \end{bmatrix} \bar{n}_t \qquad (11)$$

The boundary condition (7) can similarly be discretized:

$$n(0,t) = \int_0^{\infty} b(a)n(a,t)da = \sum_i b_i n(a_i,t), \quad \text{or}$$

$$n_{0,t+\Delta t} = \begin{bmatrix} 0 & b_1 & \cdots & b_{m-1} \end{bmatrix} \begin{bmatrix} n_{0,t} \\ n_{1,t} \\ \vdots \\ n_{m-1,t} \end{bmatrix} \qquad (12)$$

Combining (11) and (12) gives the Leslie Matrix equation:

$$\bar{n}_{t+\Delta t} = \bar{\bar{L}}\bar{n}_t \ , \quad \text{where} \quad \bar{\bar{L}} = \begin{bmatrix} 0 & b_1 & \cdots & b_{m-1} \\ v_1 & 0 & & \\ & \ddots & \ddots & \\ 0 & & v_{m-1} & 0 \end{bmatrix} \qquad (13)$$

As before, we look for solutions of the form

$$\bar{n}_t = \bar{n}_0 e^{\sigma t} = \bar{n}_0 S^t, \ where \ \sigma = \ln(S)$$

Then

$$\bar{n}_0 S^{t+\Delta t} = (\bar{\bar{L}}\bar{n}_0)S^t, \ so \ \bar{\bar{L}}\bar{n}_0 = S^{\Delta t}\bar{n}_0$$

indicating that the values $S^{\Delta t} = e^{\sigma \Delta t}$ are the eigenvalues of the Leslie Matrix.

Additional complexity can be included in the Leslie Matrix. For example, if one wanted to include copepod moulting, which results in a jump in weight class, the Leslie Matrix would have the form shown in Fig. 6.
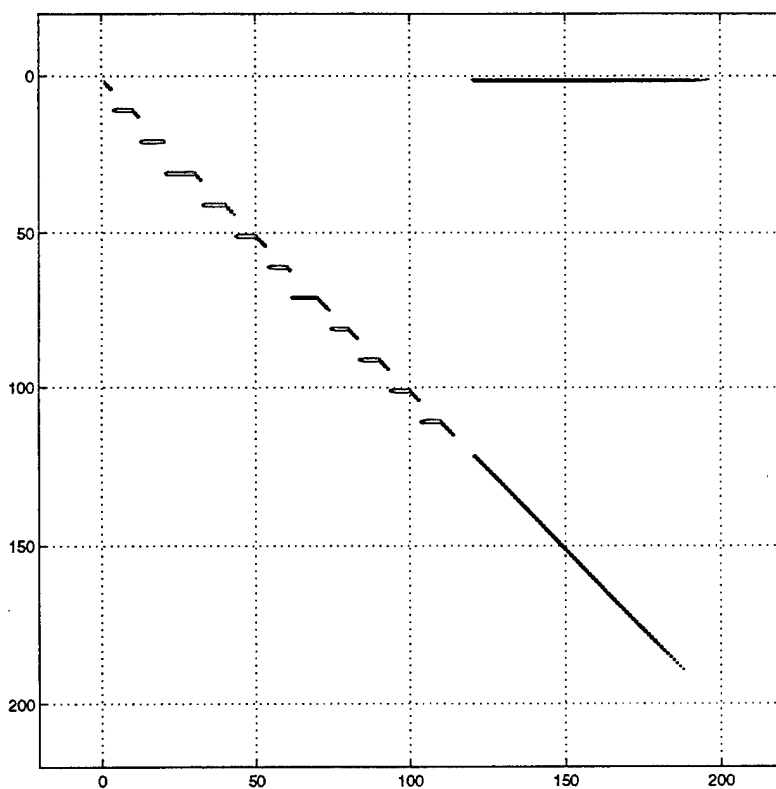
25

Figure 6: Leslie matrix for a copepod model with different age–stage classes
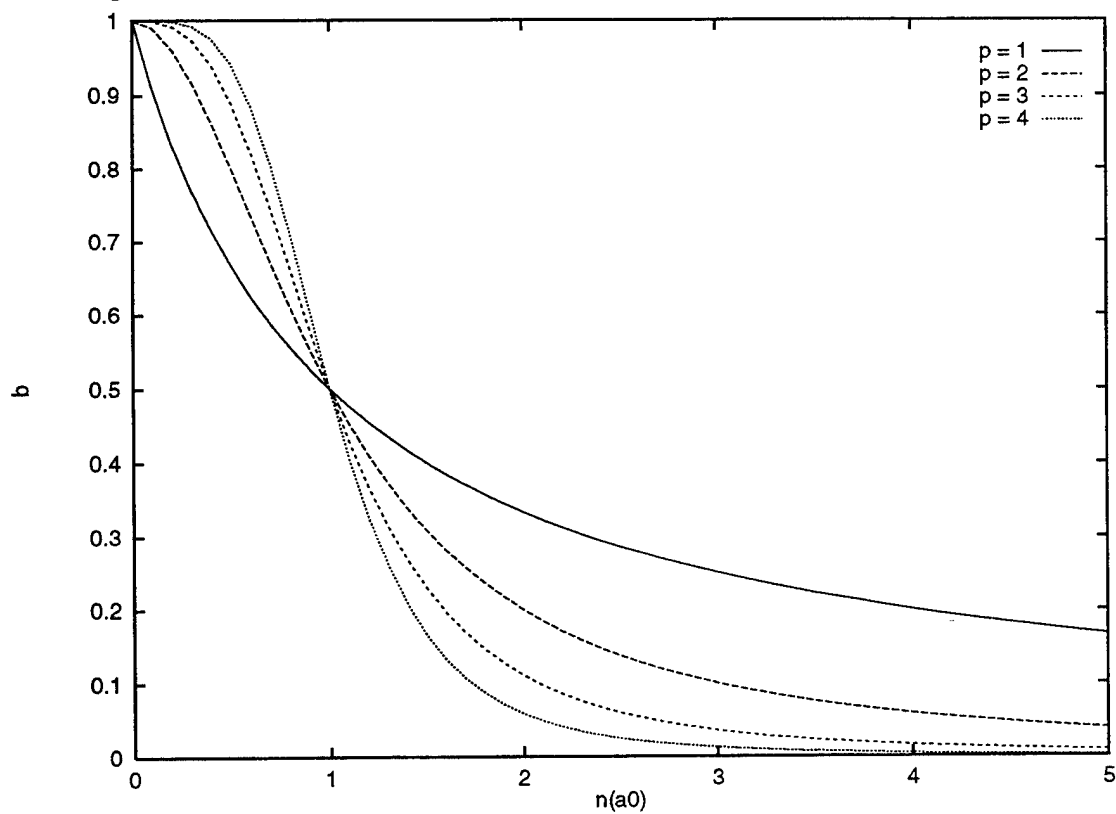

Figure 7a: Birth rate vs n curves for various power-laws $[B(x) = x^p/(1 + x^p)]$

26

Nonlinearities

Returning to equation (1)

$$\frac{\partial n}{\partial t} = -\frac{\partial}{\partial w}(gn) - \mu n$$

with the boundary condition (2) as before

$$g(w_0)n(w_0) = \int_{w_0}^{w_1} b(w', t)n(w', t)dw'$$

we see that the system will be nonlinear if g, w or b are functions of n (density dependence). For example, for territorial animals the birth rate may be dependent on both the available space (resource), R, and some function of the number of individuals:

$$b = b\left(\frac{R}{f(n)}\right)$$

a simple example of which is

$$b = B\left(\frac{R}{n(a_0)}\right)\delta(a - a_0)$$

where the density dependence is on one age only and B is some function (positive definite by definition). We choose $g = 1$ and μ and R constants. We write the solution as:

$$n(a_0, t) = n(a_0, t - a_0)e^{-\mu a_0}$$

$$n(0, t) = B\left(\frac{Re^{\mu a_0}}{n(0, t - a_0)}\right)n(0, t - a_0)$$

using (8a, 8b). Depending on the functional form of B the above solution — an iterated map — exhibits a broad range of behaviors including equilibria, limit cycles and chaos (Figure 7). Such behavior may disappear, however, when b has a wider distribution.

How can one reduce the complexity yet conserve the important characteristics? Let us rewrite the system (13) dividing the matrix $\bar{\bar{L}}$ to its subdiagonal component $\bar{\bar{L}}_0$, which is assumed to be linear, and the reproduction part that contains all the nonlinearities:

$$\bar{n}_{t+\Delta t} = \left[\bar{\bar{L}}_0 + b\left(\frac{1}{\sum n}\right)\bar{\bar{R}}\right]\bar{n}_t \tag{14}$$
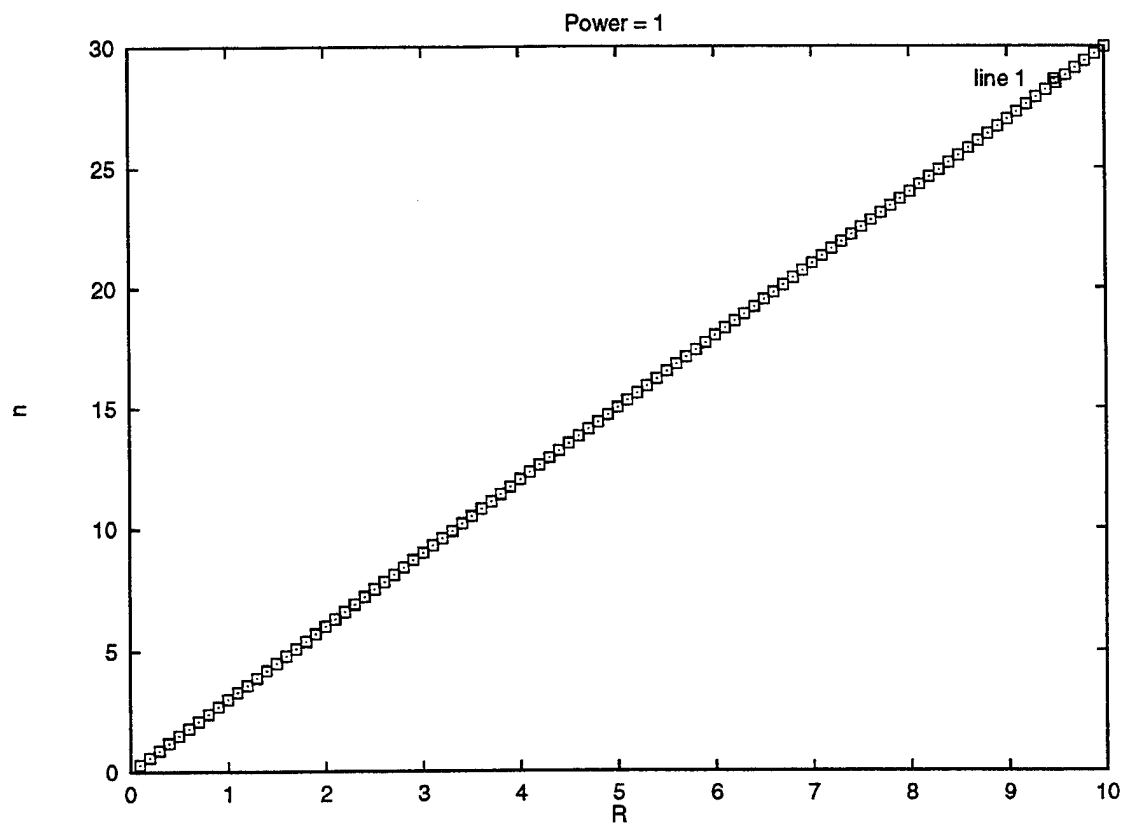
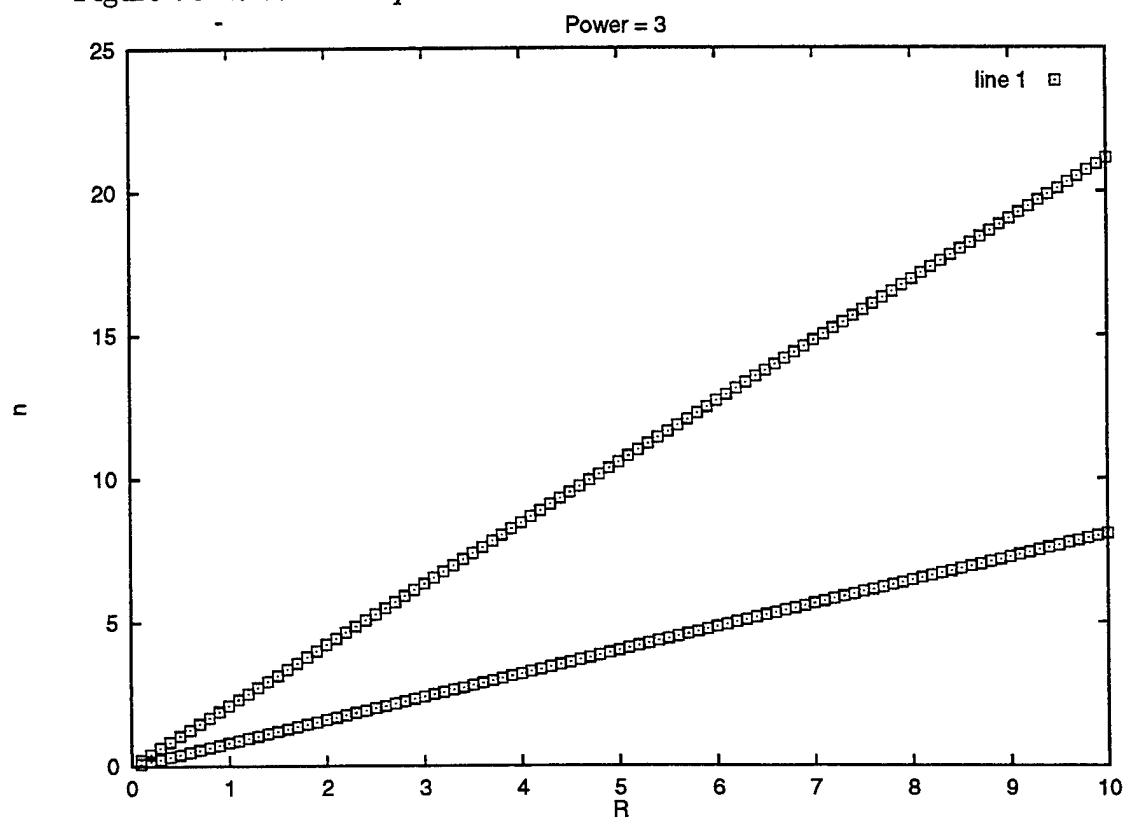Figure 7b: n vs. R for $p = 1$

Figure 7c: $p = 3$
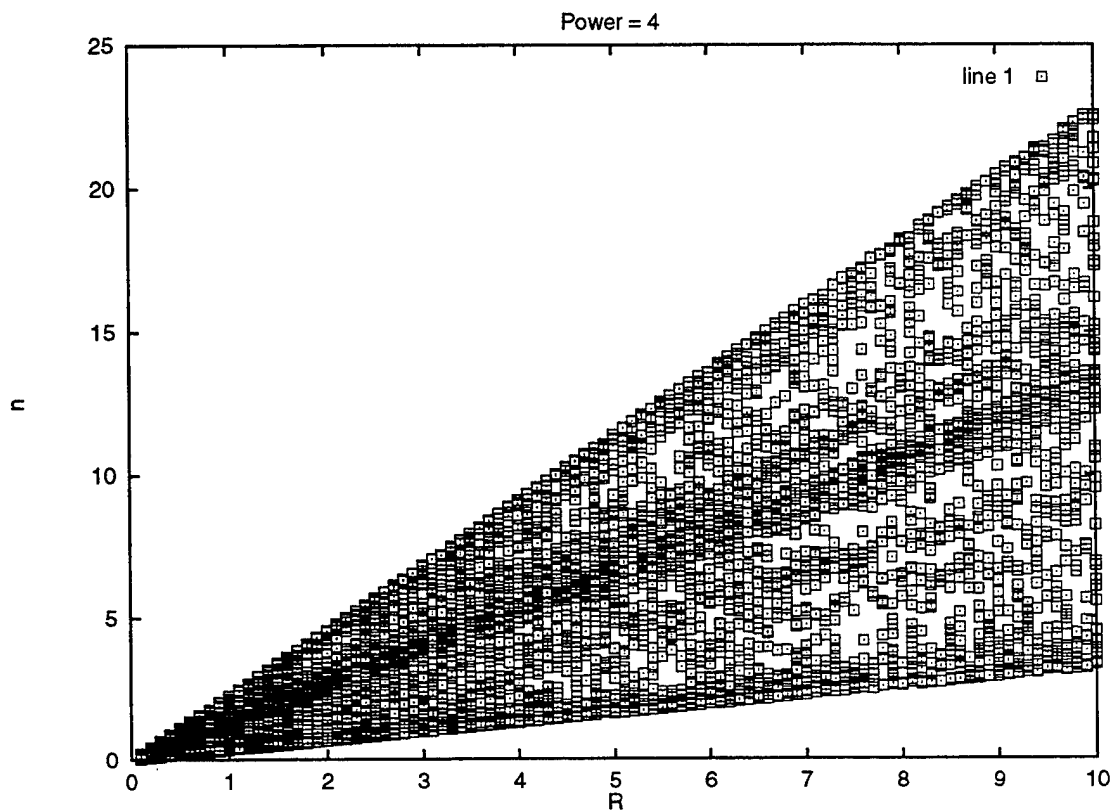
Figure 7d: n vs. R for $p = 4$
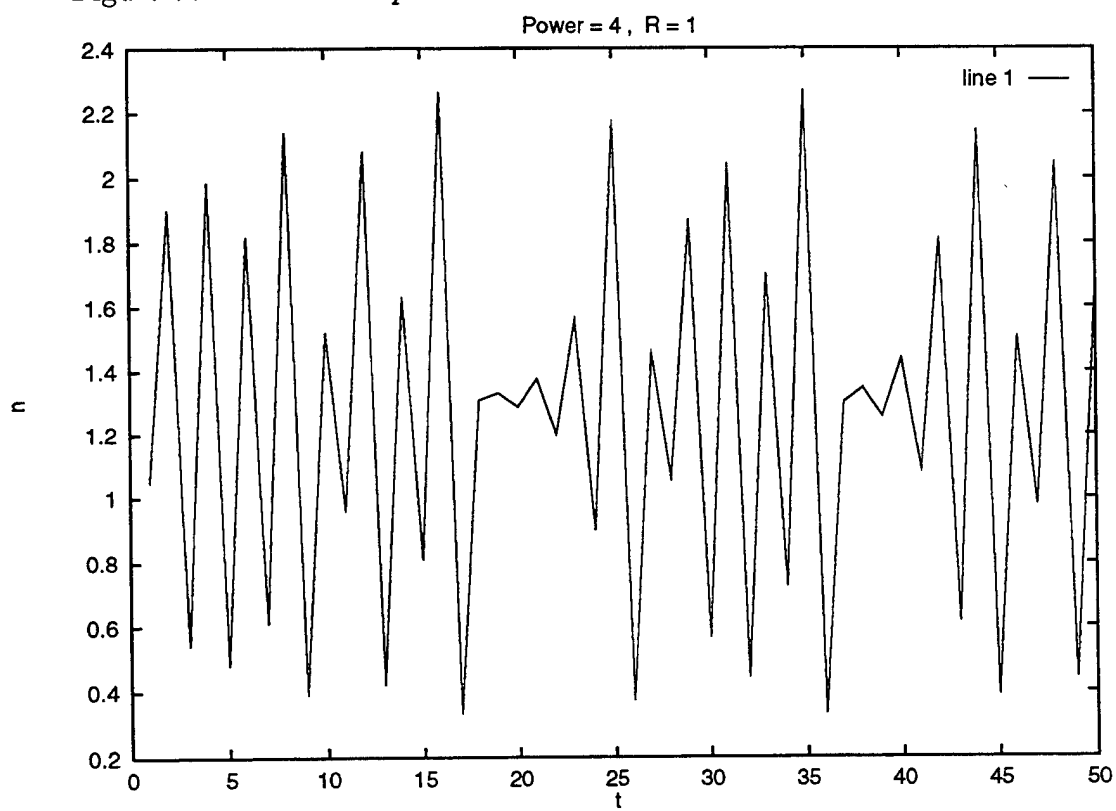
Figure 7e: n vs t for $p = 4$, $R = 1$

We've also assumed that the birth rates vary only as the total population changes and are not sensitive to the details of the age distribution. Denoting the total number of individuals by N:

$$N = \sum n \quad \text{or} \quad N = \bar{s}^T \bar{n} \quad where \quad \bar{s}^T = [1, 1, \cdots, 1]$$

and multiplying (14) by $\bar{s}^T$ we find:

$$N_{t+1} = \bar{s}^T \bar{\bar{L}}_0 \bar{n}_t + b \left(\frac{1}{N_t} \right) \bar{s}^T \bar{\bar{R}} \bar{n}_t \tag{15}$$

if one assumes that:

$$\bar{n}_t = N_t \bar{\gamma} , \quad \text{with} \quad \sum_{j=1}^{m} \gamma_j = 1$$

(15) becomes:

$$N_{t+1} = \left[\bar{s}^T \bar{\bar{L}}_0 \bar{\gamma} \right] N_t + b \left(\frac{1}{N_t} \right) \left[\bar{s}^T \bar{\bar{R}} \bar{\gamma} \right] N_t \tag{16}$$

The reduced model is again an iterated map. To calculate the coefficients, however, we must choose $\bar{\gamma}$.

One approach is to find the neutral point by varying the total population N, calculating the eigenvalues of $\bar{\bar{L}}_0 + b(1/N)\bar{\bar{R}}$, and searching for the point where the maximum eigenvalue is 1. We can use the associated eigenvector for $\bar{\gamma}$.

Another, more sophisticated, technique is to decompose the system into its normal modes by changing basis:

$$\bar{n}_t = \bar{\gamma}^{(i)} h_t^{(i)} \tag{17}$$

$h_t^{(i)}$ are the amplitudes and $\bar{\gamma}^{(i)}$ are the orthonormal eigenvectors of the system:

$$S^{(i)} \gamma^{(i)} = \left[\bar{\bar{L}}_0 + \bar{\bar{R}} \right] \gamma^{(i)}$$

Suppose that only the eigenvalue $S^{(0)}$ has magnitude greater than one and the rest are stable. Then, substitution of (17) to (14) gives:

$$h_{t+\Delta t}^{(i)} = \left(\bar{\gamma}^{(i)} \right)^{-1} \left[\bar{\bar{L}}_0 + b \left(\frac{1}{\sum n} \right) \bar{\bar{R}} \right] \bar{\gamma}^{(i)} h_t^{(i)} =$$

$$= S^{(i)} h_t^{(i)} + \left[b \left(\frac{1}{\sum n} \right) - 1 \right] \left[\left(\bar{\gamma}^{(i)} \right)^{-1} \bar{\bar{R}} \bar{\gamma}^{(j)} \right] h_t^{(j)}$$

$$= S^{(i)} h_t^{(i)} + \left[b \left(\frac{1}{h_t^{(m)} \bar{s}^T \bar{\gamma}^{(m)}} \right) - 1 \right] \bar{\bar{\Gamma}}^{(i,j)} h_t^{(j)}$$

We can estimate the amplitude of the stable modes ($h_t^{(j)}$, $j > 0$) from the steady solution

$$h_t^{(j)} \simeq \left[b \left(\frac{1}{h_t^{(0)} \bar{s}^T \bar{\gamma}^{(m)}} \right) - 1 \right] \frac{\bar{\bar{\Gamma}}^{(j,0)}}{1 - S^{(j)}} h_t^{(0)} \tag{18}$$

and use this to find an iterated map for $h_t^{(0)}$.

$$h_{t+\Delta t}^{(0)} \simeq \left[1 + (\tilde{b} - 1) \bar{\bar{\Gamma}}^{(0,j)} \right] h_t^{(0)} + \left[\tilde{b} - 1 \right] \bar{\bar{\Gamma}}^{(0,j)} h_t^{(j)}$$

with

$$\tilde{b} = b \left(\frac{1}{h_t^{(0)} \bar{s}^T \bar{\gamma}^{(0)} + h_t^{(m)} \bar{s}^T \bar{\gamma}^{(m)}} \right)$$

and $h_t^{(j)}$ coming from (18).

Lecture 3: Grunbaum June 22, 1994
Basic Formulations of Animal Aggregation and Movement

In studying animal motion in the ocean, one must keep in mind that the ocean is very dilute with respect to concentrations of organisms. The dynamics of animal motion are therefore governed largely by who encounters whom, both actively and passively. In order to provide some perspective on this problem, a hodgepodge of simple, illustrative examples will here be presented. These examples are not necessarily realistic representations of the real world; rather, they are meant to provide insight and provoke thought regarding further refinements.

Model types and descriptions of animal position and abundance be classified according to the following schemes:

1. Lagrangian (individual-based) vs. Eulerian models (continuum-based)

2. density-dependent vs. density-independent motion

Lagrangian Models:

In Lagrangian models, populations are described in terms of *individual* position or velocity (or some similar quantity); i.e., "the Lagrangian viewpoint involves identifying (marking) each individual and following the subsequent motion" (Okubo, 1980, p. 3). The basic governing equations are the equations of motion; e.g., for a group of fish:

$$m\ddot{\mathbf{X}} = \mathbf{F} = \mathbf{F}_l + \mathbf{F}_g + \mathbf{F}_d + \mathbf{F}_r + \mathbf{F}_e \tag{1}$$

where, m = mass; and $\mathbf{X}$ = position. $\mathbf{F}$ represents a collective force, which may be conceptualized in terms of its components:

$\mathbf{F_l}$ = locomotory force (tendency to maintain constant forward flying or swimming speed; this term could incorporate the effects of viscous drag or other dynamic terms)

$\mathbf{F_g}$ = grouping force (tendency to move towards or away from neighbors)

$\mathbf{F_a}$ = alignment force (tendency to match the velocity of one's neighbors)

$\mathbf{F_r}$ = random component (including both environmental and behavioral stochasticity)

$\mathbf{F_e}$ = environmental effects (including externally imposed directional tendencies, chemical gradients, physical barriers, etc.).

This representation may be simplified or elaborated upon as appropriate.

One advantage of this approach is that the modelled behavior operates at the level of the individual, as in nature, where the individual–not the school–makes the "decisions" regarding animal motion. The disadvantages are that (1) there is a practical limit to N, the number of model individuals; and (2) this approach is not amenable to analysis as are models with PDE's. Values of N are typically between 5 and 30, sometimes up to 100, and only rarely up to 10,000, whereas real schools of animals are often orders of magnitude larger. Locust swarms, for example, typically contain 10^9 individuals.

To gain some sense of how these Lagrangian models might be used, consider the example of Pacific salmon and their migration from the open ocean to their natal streams.

In the open ocean, at the beginning of their journey, the fish swim directly towards the river mouth, perhaps by compass navigation. Once in the river itself, nearing the end of their journey, the fish seem to follow primarily olfactory clues. In the coastal areas approaching the river mouth, however, it is unclear by what mechanism the fish are navigating. Here, the geography is often very complex, and the fish encounter a blend of oceanic and riverine conditions.

In exploring this general phenomenon of coastal navigation, Pasqual and Quinn (1990) considered specifically the case of sockeye salmon and their return to the Fraser River, British Columbia. One possible coastal navigation strategy for these fish might be "piloting." However, salmon returning to the Fraser River come in around both the southern and northern tips of Vancouver Island (fig. 1), although, as smolts, most of these fish exited from the river to the ocean via the northern route. It seems unlikely, then, that the return route is "chosen" based on prior experience. Ultrasonic tracking data (fig. 2) indicate that the fish do not simply trace the shoreline nor do they use tides alone to navigate. Tracking studies near northern Vancouver Island have also revealed that adult Fraser River sockeye salmon display a southeastward directional preference (the general direction of the home river). Upon encountering an obstacle, the fish were observed to swim in the opposite direction for some time before resuming their original directional tendency–perhaps as a mechanism to reduce the probability of entrapment in a bay or inlet.

These observed behaviors might result from either global preference of a compass direction or orientation to proximal information (e.g., temperature and salinity). The objective of Pasqual and Quinn (1990) was to explore the former possibility–specifically, whether the preference of a compass direction might be an effective guidance mechanism in geographically complex coastal areas. They developed a set of models, or "strategies", of sockeye salmon movement, consistent with the preference of a compass heading as a form of orientation, but with no knowledge of landmarks and no sensory contact with the target river. Since the typical transit time for a fish's entire journey is several weeks, the effects of water velocity are assumed to average out and are ignored within the model. The model fish are initially released at the head of Charlotte Strait, and their progress toward the Fraser River is modelled according to these rules of individual motion:

$$X_{t+1} = X_t + S_t \sin \phi_t \tag{2}$$

$$Y_{t+1} = Y_t + S_t \cos \phi_t \qquad (3)$$

where, S_t = swimming speed, randomly chosen from a gamma distribution taken from the data, and ϕ = direction, in degrees with respect to the north, chosen independently of S_t.

Pasqual and Quinn examined the results of four different strategies for choosing the directional angle, ϕ:

1. "uniform" strategy: null model, or random search, in which the fish direction is represented at each time step by a random deviate from a uniform distribution, taking compass values between 0 and 360°.

2. "normal" strategy: direction is added to the fish trajectory by modelling its heading at each time interval as a random deviate from a normal distribution. The mode of the distribution, μ, represents the more frequent direction in degrees with respect to the north. For each choice of μ (11 total, each constant throughout a given run), 500 individuals were run through the model, and statistics were then run on the results.

3. "inertial" strategy: a correlated random walk, keeping track of the preferred direction and the last direction:

$$\phi_{t+1} = \tan^{-1} \frac{y}{x} \qquad (4)$$

where $x = (1 - \omega) \sin(\phi_t + \Delta\phi) + \omega \sin \mu$; and $y = (1 - \omega) \cos(\phi_t + \Delta\phi) + \omega \cos \mu$. ϕ_t represents the last direction of the fish; $\Delta\phi$ represents some correction to that direction; μ represents the preferred direction; and ω represents a weighting factor.

In other words, the new direction of motion depends on the heading of the individual during the previous time step, as well as a tendency to return to the preferred direction. The strength of that tendency depends upon ω, which ranges from 0 to 1.

4. "escapist" strategy: same algorithm as the "inertial" strategy, but with the added stricture that whenever an individual encounters the shore, it uses the "uniform" strategy for a while before returning to the "inertial" strategy.

Note that the uniform and normal strategies do not incorporate memory of prior behavior, whereas the inertial and escapist strategies do.

The results of these various strategies were evaluated in terms of their success in (1) capturing the main features of sockeye salmon movement, and (2) guiding fish to the target river mouth. In both regards, the "uniform" strategy fared poorly: the simulated trajectories were very unlike the observed trajectories, and none of the 500 fish in the simulation reached the Fraser River mouth. All three directed strategies produced more realistic trajectories.

The basic results of the normal and inertial strategies are presented in figs. 3 and 4. Within the normal strategy, no combination of direction and directional precision brought more than 40% of the model salmon to the Fraser River within 28 days. For preferred headings between 135 and 145°, over 35% of the fish arrived at the goal, but this proportion declined sharply for angles outside this narrow range. The median timing for those fish reaching the goal was also strongly affected by the directional precision: more directed movement resulted in shorter migration times.

For the "inertial" strategy, higher proportions of the 500 model salmon reached the goal (fig. 4). A success rate in excess of 95% was observed for a preferred heading of 150° and intermediate directional precision. Too great a directionality resulted in inflexible movement and an inability to escape enclosed embayments, whereas too low a directionality resulted in erratic movement and little net progress towards the river mouth. A wide range of angles, for cases of higher directional precision, resulted in a quick trip to the goal (fig. 4b).

The "escapist" strategy yielded results similar to those of the "inertial" case. Figure 5 shows the spatial distribution of the 500 "escapist" fish, and overlain catch data. Note that survey catches were high for various regions of Queen Charlotte Strait, the southern coast of Johnstone Strait, and the central part of the Strait of Georgia. These patterns are all consistent with the concentrations of fish generated by the "escapist" simulation.

Another example of the application of a Lagrangian model of animal motion relates to burst-swimming fish. Many evolutionary adaptations for faster and more efficient motion in water are evident in fish: streamlined body shapes, efficient fin shapes and motion, modification of skin surface, etc. Another adaptation might be swimming "style." Weihs (1974) examined the mechanics of swimming to explore whether fish might swim more efficiently by alternating periods of accelerated motion and powerless gliding, as opposed to constant, steady swimming.

For mid- to large-sized fish, the Reynolds number (Re) is relatively high ($Re > 1000$); and for high Re, the resistance of a given rigid body is proportional to the velocity squared. For the case of a coasting fish, then, the resistance (drag) of the water is given by:

$$D = 0.5e^{AC_D}u^2 = cu^2 \tag{5}$$

where, A = surface area of the fish; C_D = coefficient of drag; and u = velocity.

Figures 6 and 7 illustrate the point that when a fish is actively propelling itself forward by undulatory motion, the drag coefficient is increased. In fig. 7, vortices can be seen shedding off the tail; the marks alongside the fish also indicate small vortices emerging out of the boundary layer–hence the increased drag. For a swimming fish, then:

$$D_s = \alpha cu^2, \qquad \alpha > 1 \tag{6}$$

A typical value for α is ~ 3.

37

Efficiency, which improves with speed, is expressed as:

$$\eta = \frac{Tu}{\text{cost to fish}} = \beta u \tag{7}$$

where, $T = $ thrust (useful work produced). Observational data indicates that efficiency is linearly related to swimming speed for a wide range of cruising speeds; hence, β is a constant.

For a coasting fish:

$$m\dot{u} = T - \alpha c u^2 \tag{8}$$

and for a swimming fish:

$$m\dot{u} = -cu^2. \tag{9}$$

Given that basic dynamic, we now have a means to address the question: What is the best strategy for crossing distance l_o within time τ–slow-plodding, steady motion (the "tortoise fish"), or alternating periods of hard swimming, coasting, hard swimming, coasting?

"Best" is here defined as that strategy which minimizes the metabolic cost of the crossing. The cost of the entire journey, E, is:

$$E = \int_0^\tau \frac{1}{\eta} Tu \, \mathrm{d}t \tag{10}$$

For the (baseline) steady-state case (i.e., crossing at constant speed, U_c):

$$U_c = \frac{l_o}{\tau} \, , \tag{11}$$

$$T_c = \alpha c U_c^2 \, , \tag{12}$$

and

$$\eta_c = \beta U_c \, . \tag{13}$$

38

The steady-state rate of working, normalized per unit energy and distance is given by:

$$\bar{E}_c = \frac{1}{\eta c} T_c U_c \tau = \frac{\alpha c}{\beta} U_c \tag{14}$$

Consider now the case of the burst-swimming fish (fig. 8). This fish starts off a speed U_i, which is lower than U_c. It produces a thrust, T, higher than that required to maintain the initial speed and accelerates according to equation (8). Upon reaching speed U_f at time t_1, the fish stops undulating and begins gliding with zero thrust, decelerating until it reaches U_i again. U_e represents the maximum speed the fish would have attained had it not begun coasting. The entire time cycle (swim–coast) is represented by $t_1 + t_2$.

As the gliding periods require no propulsive energy expenditure, the energy required per cycle (from equation (10)) is given by:

$$E_v = \int_0^\tau \frac{T}{\beta u} U \, dt \tag{15}$$

The ratio of the cost of burst swimming cost to constant swimming, R, is given by:

$$R = \frac{\bar{E}_v}{\bar{E}_c} = U_e^2 \, \frac{t_1(t_1 + t_2)}{(l_1 + l_2)^2} \tag{16}$$

where, $l_1 =$ distance travelled in the acceleration phase, and $l_2 =$ distance travelled in the coasting phase. This ratio R can serve as a quantitative criterion, showing under what conditions cyclic acceleration and gliding are advantageous to the fish (i.e., when $R > 1$). Figure 9 shows the ratio of energy required per unit length in acceleration–glide motion to that required in steady speed swimming at the same average velocity, as a function of a nondimensional average velocity, U_c/U_e. The calculations were made for various maximum speeds obtained by accelerating (solid lines). For $U_c/U_e = 1$, the two modes of swimming are identical, as the average velocity is then equal to the maximum velocity. As a result,

39

the energy required is the same, and $R = 1$. For any other (lower) average speed, there is a large range of possible acceleration–glide combinations that will save energy ($R < 1$).

The dashed line indicates the theoretical minimum of R as a function of average velocity. This minimum is 1 when the average and maximum velocities are equal, and decreases to $1/\alpha$ as the average velocity approaches zero. The shape of the dashed line indicates that the smaller the required average velocity, the greater are possible benefits of the swim–glide technique.

Eulerian Models:

Recall that, "in the Eulerian viewpoint, the flow of population individuals past a fixed point is observed" (Okubo, 1980). In these models, the conserved quantity is density, ρ, a continuum quantity. (The concept of "density" will be covered in more detail in a subsequent lecture; for the moment, think of it as a PDF of individual position, population density, etc.). The basic equation is a reaction– diffusion equation:

$$\frac{\partial p}{\partial t} = -\nabla \cdot \bar{J} + R \tag{17}$$

where, $\bar{J}$ represents a density flux. For example:

$$\frac{\partial p}{\partial t} = \nabla^2 \cdot (D\rho) - \nabla \cdot (U\rho) + R \tag{18}$$

The first term represents spatially varying diffusion; U is the local mean velocity, incorporating both behavioral and environmental components; and R represents a reaction term (i.e., a source of sink due to population dynamics). The basic question is: How does ρ behave with time?

As an example, consider the generation of a diffusion equation from a simple random walk in one dimension. Consider an animal that moves along the x-axis in discrete steps,

at discrete times. At time $t = 0$, the animal is at the origin, $x = 0$. With time, however, due to the animal's behavior, (i.e., every $\Delta t = \frac{1}{\lambda}$, the animal moves Δx to the right or left), a population of such individuals will tend to spread out. After a long enough period of time, a Gaussian distribution will be approached.

Given that $\mathrm{p}(m, N)$ is the probability of arriving at the mth gridpoint after N steps:

$$\mathrm{p}(m, N) = \frac{1}{2^N (n\ k)} \tag{19}$$

where, $(n\ k) = \frac{n!}{k!(n-k)!}$.

Using Sterling's formula, which is appropriate when N is large (i.e., for $N \gg 1$ and $N \gg m$):

$$\mathrm{p}(m, N) = \sqrt{\frac{2}{\pi N}}\ e^{\frac{-m^2}{2N^2}}\ . \tag{20}$$

Defining $x = m\Delta x$ and $N = \lambda t$:

$$\mathrm{P}(x, t)\ \mathrm{d}x = \mathrm{p}(\frac{x}{\Delta x}, \lambda t)\ \frac{\mathrm{d}x}{2\Delta x}\ . \tag{21}$$

Defining

$$D = \frac{\lambda \Delta x^2}{2}, \quad \lambda \to \infty, \Delta x \to 0 \text{ with } D \text{ constant}\ , \tag{22}$$

simple substitution yields:

$$\mathrm{P}(x, t) = \frac{1}{\sqrt{4\pi D t}}\ e^{\frac{-x^2}{4Dt}}\ . \tag{23}$$

In this very simple case, then, we have now related the individual behavior described above to a probability, $\mathrm{P}(x, t)$, which is continuous.

To get a description of the dynamics of overall behavior (i.e., starting with a description of individual behavior and now moving towards a global description of the population), observe that this distribution obeys a particular PDE (the diffusion equation):

$$\frac{\partial P}{\partial t} = D\frac{\partial^2 P}{\partial x^2}\ . \tag{24}$$

41

An important point here is that this PDE is a lucky fact which is observed *after* the distribution above is obtained. This fortunate occurrence has predisposed people to expect that there will always be a suitable PDE found at the end, but not so! (The same is true in the case of fluid dynamics).

Using this (eqn 24), we can now perform statistical analyses regarding how the population might be expected to behave. We might, for example, calculate the mean-square displacement:

$$< x^2 >= \int_{-\infty}^{\infty} x^2 P(x,t) \, \mathrm{d}x \qquad (25)$$

with the result that

$$< x^2 >= 2Dt \; . \qquad (26)$$

This final result is important, because in sampling populations we often can obtain only statistical information (as in eqn. 26), rather than continuous rate information (as in eqn. 24).

Levin and Segel (1976) apply the concept of this diffusive tendency to the study of plankton patchiness in the ocean. Letting P represent the population of phytoplankton, and H, the population of herbivores:

$$\frac{\partial P}{\partial t} = aP + eP^2 - bPH + \mu \frac{\partial^2 P}{\partial x^2} \qquad (27)$$

$$\frac{\partial H}{\partial t} = cPH - dH^2 + \nu \frac{\partial^2 H}{\partial x^2} \qquad (28)$$

where, aP=linear growth; eP^2=nonlinear growth rate; μ = phytoplankton diffusivity; cPH=consumption term; and ν=herbivore diffusivity. Equation 27 is applicable to a post-cropped population of phytoplankton.

Note that there are two important properties incorporated into these expressions: (1) an "autocatalytic" effect in phytoplankton density (i.e., nonlinear growth, with phytoplankton growing more rapidly when more individuals are present), attributable to a

reduced efficiency of herbivory as phytoplankton density increases; and (2) differential dispersal rates, since herbivores are motile animals (i.e., $\nu > \mu$).

This system is seen to have a spatially uniform stable equilibrium. Setting all derivative terms to 0:

$$\bar{P} = \frac{ad}{bc - ed} \qquad (29)$$

$$\bar{H} = \frac{ac}{bc - ed} \qquad (30)$$

provided that $bc - ed > 0$ and $c > e$. $\bar{P}$ and $\bar{H}$ represent equilibrium values.

These basic dynamics of the system, considered in a spatial context, are somewhat counterintuitive, as diffusion is often thought of as a "spreading-out," stabilizing, homogenizing process. But can it create pattern by destabilizing an otherwise stable interaction? Introducing a small-amplitude perturbation to the system:

$$P = \bar{P} + P' \cos(kx)e^{st} \qquad (31)$$

$$H = \bar{H} + H' \cos(kx)e^{st} \qquad (32)$$

and assuming $P', H' \ll \bar{P}, \bar{H}$ yields:

$$P^2 = (\bar{P}^2 + 2\bar{P}P + P'^2) \qquad (33)$$

Since $\bar{P} \gg P'$, the P'^2 term goes to zero, as does the $\bar{P}^2$ term.

At issue is the question of whether s is ever greater than 0 in a case where the non-spatial case is stable. Defining

$$R = \frac{\nu}{\mu}, \qquad (35)$$

and

$$R_c = \frac{1}{(\sqrt{\frac{b}{d}} - \sqrt{\frac{b}{d} - \frac{e}{c}})^2}, \qquad (36)$$

the uniform state is found to be stable for $R < R_c$, but unstable for $R > R_c$ (i.e., nonuniform perturbations grow).

For R only slightly greater than R_c, the uniform state is replaced by a new steady state in which phytoplankton and herbivores are more concentrated in certain regions (fig. 10). The pattern is maintained because the herbivores have a higher diffusivity than do the phytoplankton (i.e., $\nu > \mu$) and thus tend to "leak out" into phytoplankton-poor areas, enabling the phytoplankton to persist.

References:

Levin, Simon A. and Segel, Lee A. (1976) Hypothesis for origin of planktonic patchiness, *Nature*, 259, 659.

Okubo, A. (1980) *Diffusion and Ecological Problems: Mathematical Models.* Springer-Verlag.

Pascual, Miguel A. and Thomas P. Quinn (1991) Evaluation of alternative models of the coastal migration of adult Fraser River sockeye salmon (*Oncorhynchus nerka*), *Canadian Journal of Fishery and Aquatic Sciences*, 48, 799-809.

Weihs, D. (1974) Energetic advantages of burst swimming of fish, *Journal of Theoretical Biology*, 48, 215-229.

Figures:

1. Northern route to the Fraser River, British Columbia. (Shaded = land; white = water). 1 = Queen Charlotte Strait; 2 = Johnstone Strait; 3 = Strait of Georgia. **D** = departure line for the model fish; **G** = goal. ¿From Pascual and Quinn (1990).

2. Trajectories of six sockeye salmon as revealed by ultrasonic tracking. Closed circles represent release points. From Pascual and Quinn (1990).

3. Contours of (a) percent success in reaching the goal, and (b) median timing in days for 500 simulated fish under the "normal" movement strategy for different values of preferred heading (μ) and directional precision (= coefficient of variation of normal heading distribution). From Pascual and Quinn (1990).

4. Contours of (a) percent success in reaching the goal, and (b) median timing in days for 500 simulated fish under the "inertial" movement strategy for different values of preferred heading (μ) and directional precision (ω = weight to the preferred direction). From Pascual and Quinn (1990).

5. Spatial distribution generated by 500 simulated fish for the "escapist" movement strategy for a preferred heading (μ) of 125° (shaded areas) and observed catch rates (closed circles). **D** represents the departure line for the model fish; **G** represents their goal. From Pascual and Quinn (1990).

6. Visualization of flow over *Trachurus mediterraneus politicus* during movement by inertia and passive breathing. Synchronous film frames (plan and side views). $V = 0.52$ m/s, $Re = 7.1$ x 10^4. From Aleev (   ).

7. Visualization of fish movement. Horizontal line indicates point of boundary-layer transition from laminar (L) to turbulent (T) flow. Dashes mark individual microvortices in the turbulent boundary layer, and bent arrows indicate sense of vorticity in the wake. From Aleev (   ).

8. Velocity as a function of time for burst-swimming fish. U_i = initial velocity; U_f = highest velocity attained during cycle; and U_e = maximum sustained velocity (i.e., velocity the fish would have attained had it not begun coasting). U_c = velocity of the constant-speed fish.

9. R, the ratio of energy per unit distance required for acceleration–gliding motion to that required for steady speed swimming, versus the average velocity (normalized by the maximum sustained speed, U_e). U_f = highest velocity attained during cycle. ¿From Weihs (1974).

10. Creation and maintenance of phytoplankton patchiness by diffusive instability.
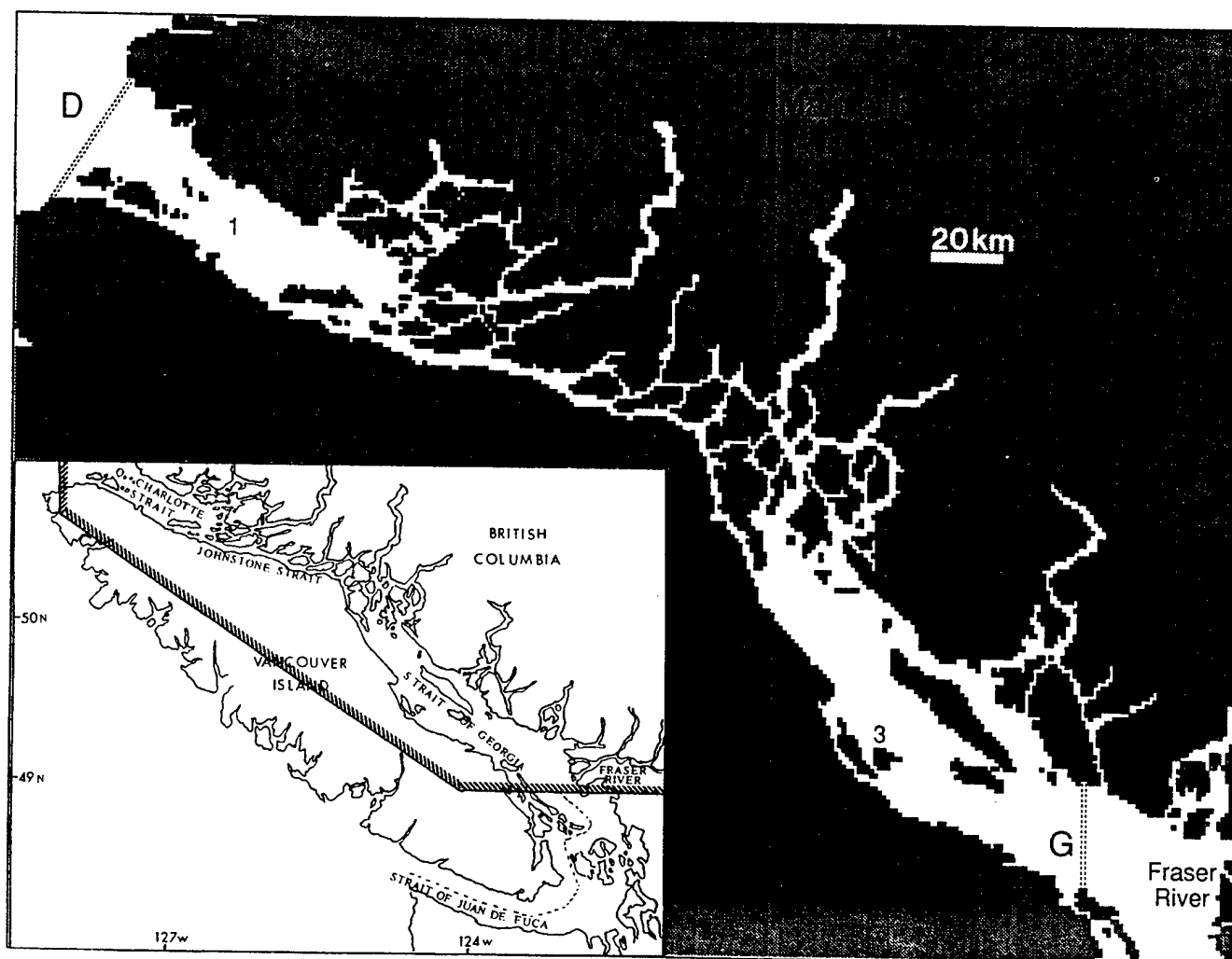
3.1

FIG. 1. Northern route to the Fraser River and the graphical representation of the binary space matrix used in the simulations (shaded = land, white = water, D = departure line, G = goal, 1 = Queen Charlotte Strait, 2 = Johnstone Strait, 3 = Strait of Georgia).
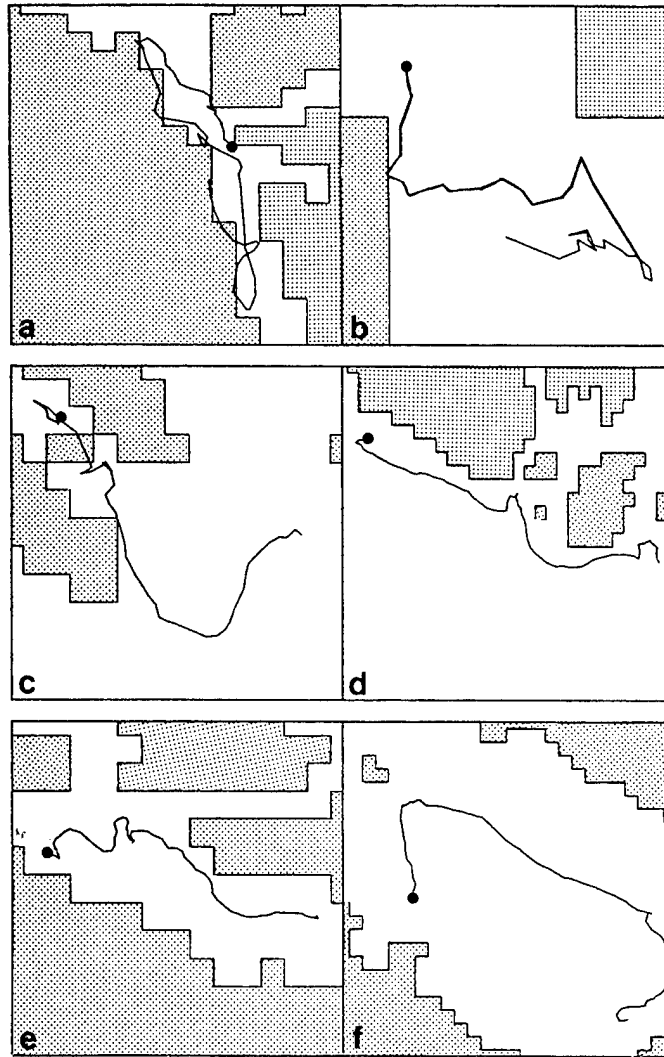
3.2

FIG. 2. Trajectories of six sockeye salmon as revealed by ultrasonic tracking (from Quinn and terHart 1987; Quinn et al. 1989). ● = release point.
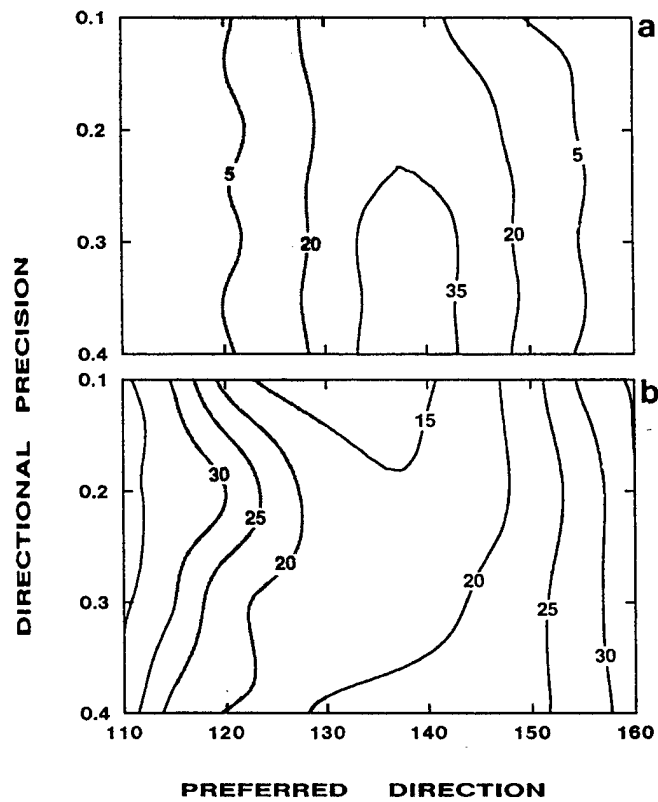
3.3

FIG. 5. Contour graphs of (a) percent of success in reaching the goal and (b) median timing in days for 500 simulated fish under the *normal* movement strategy for different values of preferred heading (μ) and directional precision (= CV of normal heading distribution). See text for interpretation of shaded areas.
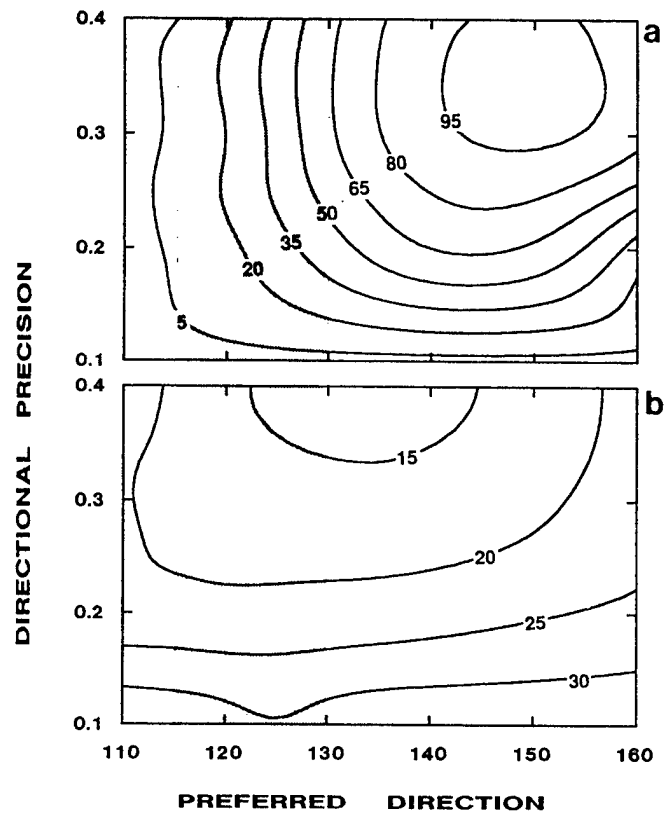
49

3.4

FIG. 7. Contour graphs of (a) percent of success in reaching the goal and (b) median timing in days for 500 simulated fish under the *escapist* movement strategy for different values of preferred heading (μ) and directional precision (ω = weight to the preferred direction). See text for interpretation of shaded areas.
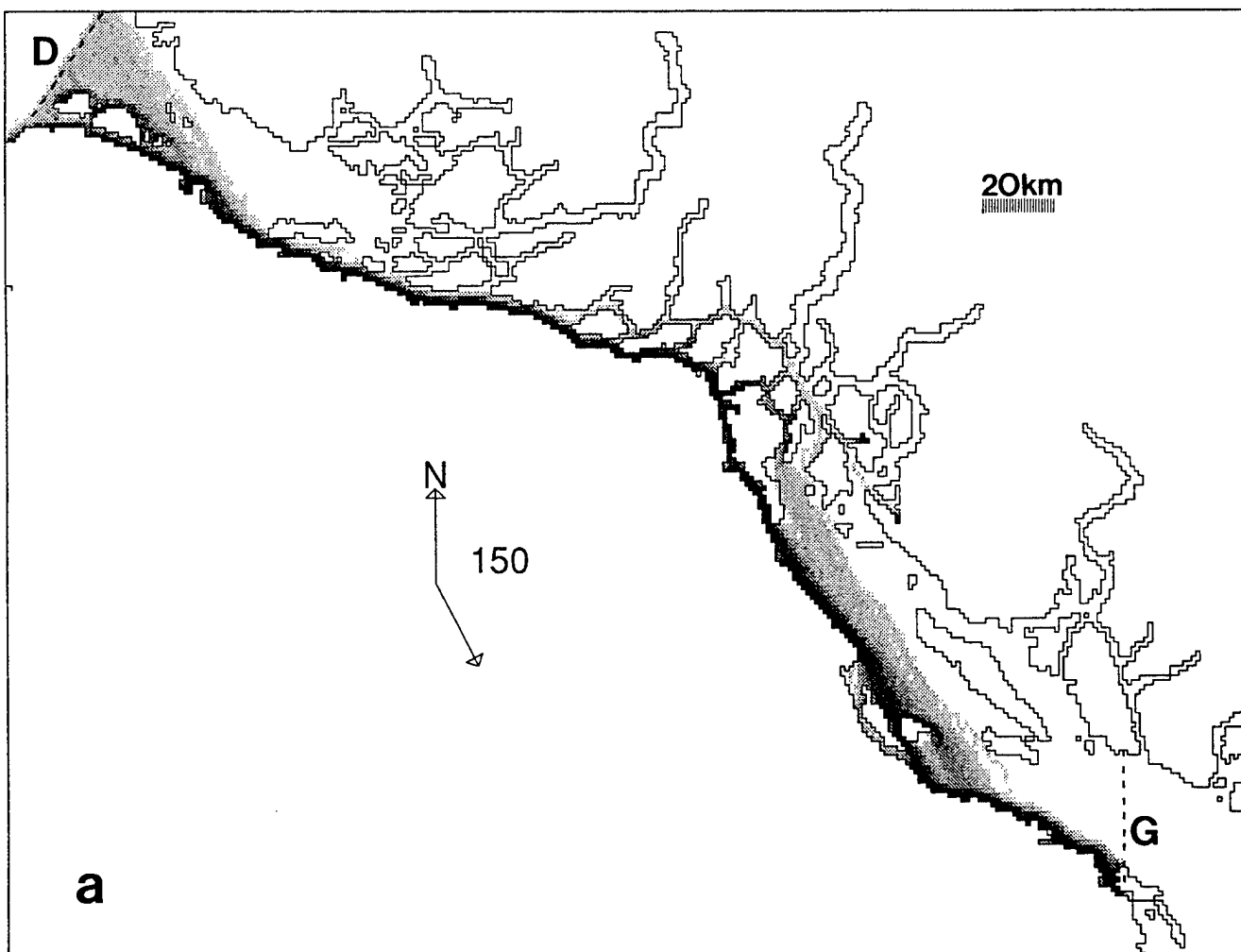
3.5

FIG. 8. Spatial distributions generated by 500 simulated fish for the *escapist* movement strategy for preferred headings (μ) of (a) 150° and (b) 120° (D = departure line, G = goal). Intensities in the shaded areas indicate relative abundances (logarithm of the number of times a cell was visited by a fish). (*Fig. 8 concluded next page*)

3.6

Figure 74. Visualization of flow over *Trachurus mediterraneus ponticus* Aleev during movement by inertia and passive breathing. Synchronous film frames (plan and side views) from filming with two cameras. $L_a = 13.6$ cm, $V = 0.52$ m/s, $Re = 7.1 \times 10^4$. See text.
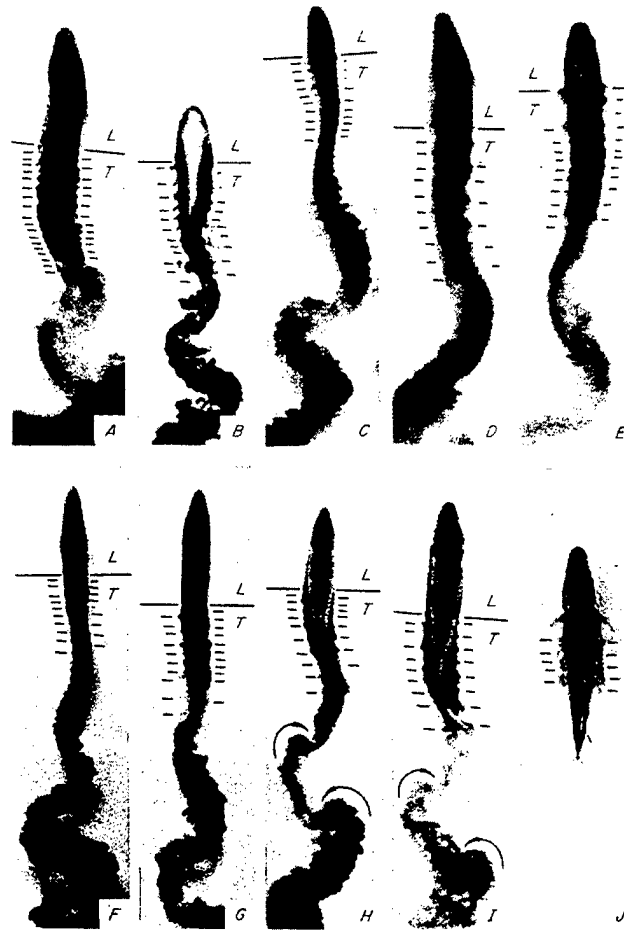
Figure 72. Visualization of fish movement. *A, Salmo gairdneri irideus* Gibb., $L_a = 16.3$ cm, $V = 0.75$ m/s, $Re = 1.2 \times 10^5$; *B, Carassius auratus* (L.), $L_a = 12.2$ cm, $V = 0.29$ m/s, $Re = 3.5 \times 10^4$; *C, Mugil auratus* Risso, $L_a = 12.0$ cm, $V = 0.95$ m/s, $Re = 1.1 \times 10^5$; *D, Mugil saliens* Risso, $L_a = 19.0$ cm, $V = 0.76$ m/s, $Re = 1.4 \times 10^5$; *E, Mugil auratus* Risso $L_a = 18.0$ cm, $V = 0.48$ m/s, $Re = 8.7 \times 10^4$; *F, Chalcalburnus chalcoides* (Gül.) $L_a = 13.5$ cm, $V = 0.84$ m/s, $Re = 1.1 \times 10^5$; *G, Spicara smaris* (L.), $L_a = 15.9$ cm, $V = 0.46$ m/s, $Re = 7.4 \times 10^4$; *H, Diplodus annularis* (L.), $L_a = 16.4$ cm, $V = 2.00$ m/s, $Re = 3.3 \times 10^5$; *I, Diplodus annularis* (L.), $L_a = 16.4$ cm, $V = 0.76$ m/s, $Re = 1.2 \times 10^5$; *J, Diplodus annularis* (L.), $L_a = 16.4$ cm, $V = 0.36$ m/s, $Re = 4.5 \times 10^4$. Horizontal line indicates point of laminar (L) to turbulent (T) boundary-layer transition; dashes mark individual microvortices in the turbulent boundary layer; bent arrows indicate sense of vorticity in the wake.

53

3-8
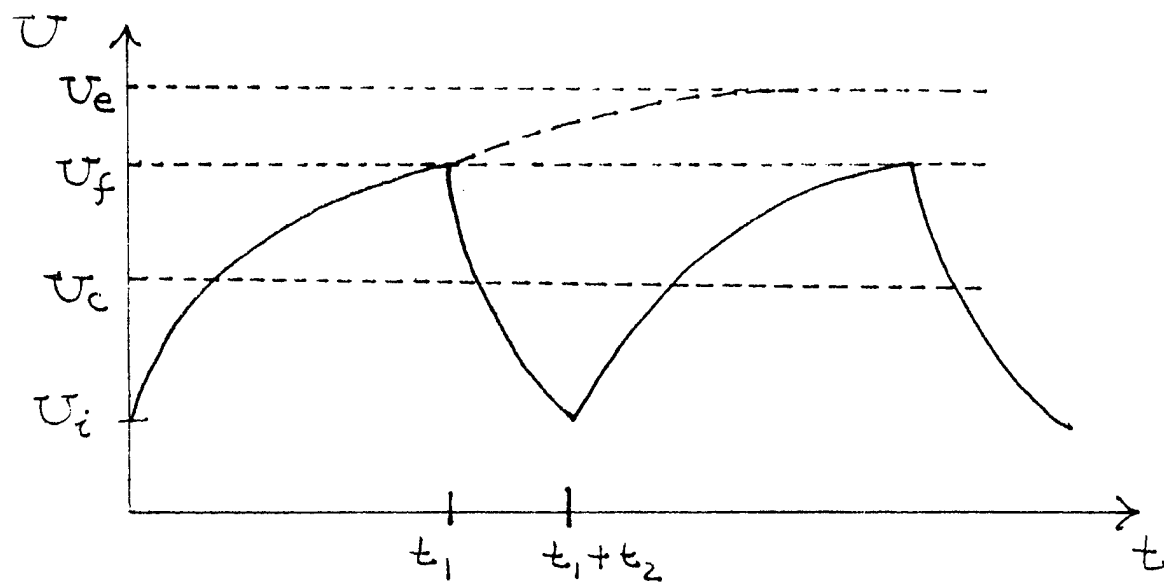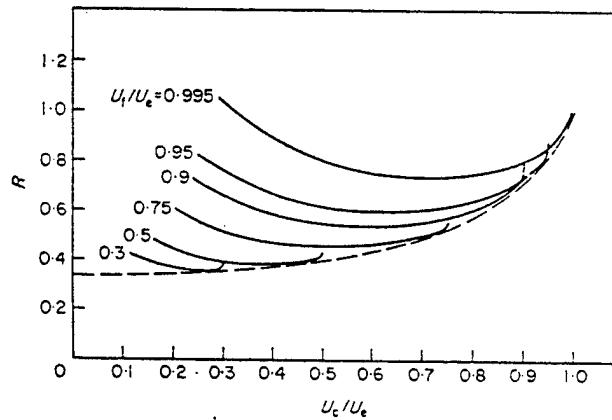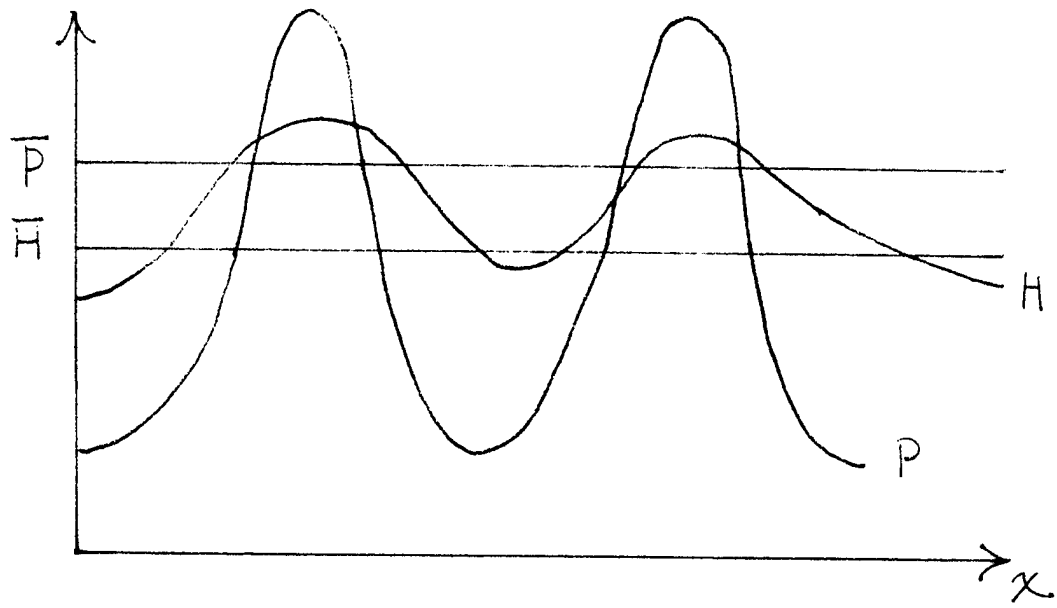
3.9

BURST SWIMMING OF FISH

FIG. 1. The ratio of energy per unit distance traversed required in acceleration–gliding motion, to that required for steady speed swimming R, versus the average velocity (normalized by the maximum sustained speed U_e). U_t = highest velocity attained during cycle. The dotted line shows the greatest possible saving.

55

3.10

Lecture #4 - Density-dependent Models of Social Aggregation[1]

Lagrangian Models

Schooling behavior of fish is one problem in density dependent social aggregation studies amenable to Lagrangian description. Fish are observed to travel in groups wherein individuals maintain a characteristic distance from other individuals in the group and they match both the speed and direction of other individuals within the group. As shifts occur, individuals change their position and velocity to maintain their characteristic distance from other individuals and to align their velocity with other individuals.

While many social aggregation models are designed to provide heuristic insight rather than quantitative prediction, simplification is still necessary for both classes of models due to the complexity of factors driving aggregation behavior. While simplifying assumptions are necessary in order to produce models of schooling behavior, there are biological motivations for many of the assumptions commonly used. For example, defining a series of threshold distances which control the directional behavior of modeled individuals is an approach firmly rooted in the biological characteristics of fish. Fish have more than one sensory system with which they can detect their neighbors and these different apparatuses may induce different behavioral responses to neighbors. The lateral line system of fish is sensitive to pressure fields and could be responsible for influencing individuals to swim away from each other in order to maintain a certain minimum distance between them and their neighbors in the school. This asserted influence of the lateral line system is supported by experimental evidence that shows that fish whose lateral line has been severed do not tend to maintain as large a minimum distance from other members of their school as do intact individuals. In contrast, the tendency for schooling individuals to swim toward each other, if they are beyond a certain threshold distance from other school members, may be visually cued. Visual control of maximum threshold distance might tend to allow individuals to rejoin their school as long as the school is within their field of visual perception.

---

[1]Notes prepared by Cathryn Rhodes and Richard Matear

One formulation for modelling schooling behavior based on fish changing their swimming direction was investigated by Aoki (1982). To simplify the problem of modelling schooling behavior, Aoki created a discrete time model in which: 1) decisions at each time step are independent of previous time steps, 2) movements are confined to 2 dimensions and occur on the horizontal plane, 3) speed and direction are independent stochastic variables, 4) interactions between individuals cause only directional changes; speed is independent of other individuals and is chosen from a Γ distribution, and 5) initially, individuals are distributed randomly within a square and their directions are uniformly distributed from 0° to 360°. Individuals are considered to have concentric zones surrounding them, defined by their radial distance (Figure 1). The zones can be conceptualized as distances at which the focal fish is either "comfortable" or "uncomfortable" with its distance from the other members of the school. If the focal fish is too close or too far away from the school, it is judged to be "uncomfortable" with its distance from the school and swims away from or toward the school until it is "comfortable" with its position relative to the school. However, if the other individuals in the school are a "comfortable" distance from the focal fish, that fish will align its swimming direction with that of the other members of the school. The focal fish is considered to make these choices based on the following decision rules where r is the radial distance from the focal to the jth individual:

if

$r_2 \leq r_j \langle r_3$	swim toward neighbor
$r_1 \leq r_j \langle r_2$	swim parallel to (align with) neighbor
$r_j \langle r_1$	swim away from neighbor.

The probability density for the direction of movement of the ith individual is given by:

$$p_i(\theta) = \sum_j W_j \frac{1}{S_j \sqrt{2\pi}} e^{\frac{-(\theta - M_j)}{2 S_j^2}} \tag{1}$$

where:

j=number of neighbors in the school

θ = the angle describing the direction of the ith individual

M_j=the change of direction induced by the jth individual

W_j=weighting factor for the influence of the jth individual

S_j=distance between the ith and jth individual

This formulation will yield a probability distribution for the angle θ that looks rather complex; however, this distribution is really a composite of many Gaussian distributions (Figure 2). In this model, it is the M_j's that incorporate the variability in behavior between the individual members of the school.

Eulerian Models

Density dependence can also be modeled in the Eulerian sense. The question can be asked, "How do cells interact to produce a clumped distribution?" One method for creating Eulerian models is to consider populations of organisms in analogy with physical systems and to examine cell population densities using the cell potential approach (Cohen and Murray 1981). If we consider a gradient in potential μ representing a flux J that is proportional to $\nabla\mu$, we can think of the potential as the work done by changing the state by a small amount. If we have a spatial distribution of cell density, $n(x,t)$, and an associated internal energy per unit volume, $e(n)$, of the developing spatial pattern, we may write

$$E[n] = \int_V e(n)dx \qquad (2)$$

where,

$E[n]$=the total energy

$e(n)$=the total energy per unit volume.

The work done in changing states by a small amount--δn--is δE; from this we can define the cell potential, $\mu(n)$, as the variational derivative $\delta E/\delta n$,

$$\mu(n) = \frac{\delta E}{\delta n} = e'(n) \qquad (3)$$

A flux J is produced proportional to $\nabla\mu$, so

$$J = -D\nabla\mu(n) \qquad (4)$$

Therefore the continuity equation for n is

$$\frac{\delta n}{\delta t} = -\nabla \cdot \mathbf{J} = \nabla \cdot [D \nabla \mu(n)] = \nabla \cdot [D e''(n) \nabla n] = \nabla \cdot [D^*(n) \nabla n] \tag{5}$$

where

$$D^* = D e''(n) \tag{6}$$

In the classical diffusion case where diffusion is considered to be constant, the internal energy $e(n) = n^2/2$. In this situation, $\mu(n) = n$, and $D^* = D$ and if D is a function of x, t, and n, (5) becomes

$$\frac{\partial n}{\partial t} = \nabla \cdot [D^*(x, t, n) \nabla n] \tag{7}$$

In this equation, energy is proportional only to cell density. In an environment where there may be spatial heterogeneity, internal energy associated with a cell distribution might depend on the gradient in cell density.

If we take a more realistic energy function,

$$E[n] = \int_V [e(n) + \frac{k}{2}(\nabla n)^2] dx \tag{8}$$

we can use the Landau-Ginzburg free energy form

$$e(n) = \frac{an^2}{2} + \frac{bn^4}{4} \tag{9}$$

from which we can obtain

$$\mu = an + bn^3 - k\nabla^2 n \tag{10}$$

and

$$\frac{\partial n}{\partial t} = -\nabla(-D\nabla(an + bn^3 - k\nabla^2 n)) \tag{11}$$

This can also be written as

$$\frac{\partial n}{\partial t} = Da\nabla^2 n + Db\nabla^2 n^3 - Dk\nabla^4 n \qquad (12)$$

where

$Da\nabla^2 n \equiv$ diffusion

$Db\nabla^2 n^3 \equiv$ higher order dispersive term

$Dk\nabla^4 n \equiv$ aggregation term

So far, this description is only phenomenological. However, certain morphogenetic characteristics of motile cells could be described by this model. Attaching biological meaning to these equations depends on identifying "internal energy" with specific individual behaviors and positions.

Another Eulerian approach for modeling density dependent population effects is to use an integral equation for the spatial dependence of density, ρ. This approach has been used by Kawasaki(1978) and Alt(1985). If we take

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2} - \frac{\partial}{\partial x}(\rho \cdot U) \qquad (13)$$

where the first term is random walk diffusion and the second term contains the integral density dependence as

$$U = \int_{-\infty}^{\infty} W(x'-x)\rho(x')\,dx \qquad (14)$$

then $W(x)$ is the response to density of neighbors as a function of distance. We can diagram W as a function of x (Figure 3). Near the origin, for example, there might be a zone of repulsion, which changes to a zone of attraction as distance between individuals increases.

It would be more analytically and computationally convenient to have a PDE rather than an integral form. Therefore, we shall examine PDE approximations to the PIDE

form in Figure 3. We will compare the linear stability properties of the integral form, with the understanding that appropriate stability behaviors are a necessary property of the simplified equation.

Expand the animal density, ρ, by considering a sinusodial perturbation ρ' to the mean density ρ as follows

$$\rho(x,t) = \overline{\rho} + \rho' \sin(kx) e^{st} \qquad (15)$$

Substituting this equation into the integral partial differential equation and computing the sine tranform of the new equation, one obtains

$$S = -k^2 D + \overline{k\rho} W(k), \qquad W(k) = \int_{-\infty}^{\infty} W(x) \sin(kx)\, dx \qquad (16)$$

As an example, by defining the convolution function that describes the animal behavior as

$$W(x) = x e^{-x^2} \qquad (17)$$

one can obtain an explicit equation for the sine transform of this function, which is

$$W(k) = \frac{1}{2}\sqrt{\pi} k e^{-\frac{k^2}{4}} \qquad (18)$$

As shown in Figure 4, the sine transform term in the integral formulation (Equation (16)) approaches zero as wavenumber k becomes large. Thus, short wavelength disturbances are dissipated quickly by diffusion. Using a Taylor series expansion of convolution integral, one can writh the nth order approximation to an individual velocity as

$$U_n = \sum_{i=0}^{n} \frac{\partial^2 \rho}{\partial x^2} \int_{-\infty}^{\infty} \frac{x^i}{i!} W(x)\, dx \ . \qquad (19)$$

Subsituting this equation into the integral partial differential equation, and again calculating the growth rate, one obtains the following equation

$$S=-k^2 D=\overline{k\rho}\,W(k), \qquad W(k)=\sum_{k=0}^{n}\frac{k^i}{i!}\frac{\partial^i W(k)}{\partial k^i} \tag{20}$$

A polynomial approximation, (20), to the true growth rate (16) shows divergent behavior for large wavenumber (Fig. 4). Thus a PDE approximation derived from a Taylor series, as in (19), would be expected to have erroneously growing modes either in the actively aggregating or actively disaggregating case, depending on the order of the approximation.

Generalized Social Behaviour

For Eulerian models of density-dependent interactions, the flux density involves some sort of term representing the average "decision" made by individuals at a particular time and place. What follows here is a simple demonstration that the density above does not contain enough information to specify the expected flux, even if the underlying behavior is known. Thus additional assumptions are needed which specify more fully when and where animals are actually found. Figure 5 shows a schematic of the different zones sensed by a fish. If an individual fish senses another fish in one of these zones the fish will react to this fish. For example, a fish may try to move toward fish in subdomains 1 to 4 while trying to maintain its current distance from fish in domains 5 to 8.

By defining the number of neighbors in subdomain d_i as n_i, the behavioral response of an individual is determined by sensing the number of other individuals in its vicinity. The behavioral response is described as the movement decision function, $f(x,t)=f\,(n_1,\,n_2,\,...,\,n_M)$. The expected movement of an individual is calculated from the expectation of the function at a given time and location,Using a Taylor series

$$\Phi(x,t)=E\{f(x,t)\} \tag{21}$$

expansion, the movement decision function can be expanded

$$f=f(v_1,v_2,...,v_M)=f(\rho_1,\rho_2,...,\rho_M)+\sum_{i=1}^{M}(v_i-\rho_i)\frac{\partial f}{\partial \rho_i}+\frac{1}{2}\sum_{i=1}^{M}(v_i-\rho_i)(v_j-\rho_j)\frac{\partial^2 f}{\partial \rho_i \partial \rho_j} \tag{22}$$

where $\rho_i=\{E v_i\}$, is the average in the sample domain. From this expansion, the expected movement of an individual is

63

$$\Phi = e\{f\} = f(\rho_1, \rho_2, ..., \rho_M) + (0) + \frac{1}{2} \sum_{i=1}^{M} E\{(v_i - \rho_i)(v_j - \rho_j) \frac{\partial^2 f}{\partial \rho_i \partial \rho_j} = ... \ . \tag{23}$$

This leads to a closure problem which requires higher moments of the density distribution to evaluate the expected movement of an individual. If one assumes a Poisson distribution in the animal density, the second and higher order moments are calculable and the expected movement animals at a given time and location can be determined.

Integral Equation for Density-Dependent Swarming

Lagrangian Model

The equation of motion of an individual can be written as

$$\frac{\partial^2 x}{\partial t} + \frac{\partial x}{\partial t} = F(t) + U(x,t) \tag{24}$$

where the function $U(x,t)$ demotes an advective veloctiy and the function $F(t)$ describes the behavioral forces acting on an individual. The behavioral forces can be separated into two terms

$$F(t) = F_r(t) + F_s(t) \tag{25}$$

The first term of the behavioral forces, $F_r(t)$, describes random forces acting on an individual which are chosen at an average rate λ_r, from a Gaussian distribution of zero mean and standard deviation σ^2. The second term of the behavioral forces, $F_s(t)$, defines "social" forces acting on an individual. These forces are chosen at an average rate λ_s with constant magnitude α, and direction determined by social response to its neighbors.

Figure 6 shows schematically how the density of individuals over a unit sensing range is used to determine the social response of an individual. This response is as follows: 1) count the number of individual within sensing range to estimate local density; 2) estimate direction of gradient, i.e. count which side has more neighbors, and move down gradient if above "target density", and up gradient if below.

In one dimension, the mathematical representation of this response can be described by the following equation

$$F_s = \alpha g(v_1, v_2, \mu), \quad g = sgn[2\mu - 1 - v_1 - v_2)(v_2 - v_1)] \tag{26}$$

where at a particular social response, v_1 is the number of neighbors on the left hand side and v_2 is the number of neighbors on the right hand side, and μ is the target situation. The individual will move in the direction of the target density μ.

Eulerian Model

An Eulerian representation of the Lagrangian model can be formulated by using the following reparameterization.

$$\lambda = \lambda_r + \lambda_s; \quad \gamma = \frac{\alpha \lambda_s}{\lambda}, \quad \Delta x_s = \frac{\alpha}{\lambda}. \tag{27}$$

In this reparameterisation, D defines the diffusitvity, γ defines the characteristics aggregate veleocity, λ is the total rate of choosing new forces, and Δx_s is the characteristic steplength for social responses. A further assumption to deriving the Eulerian model is that the individual's environment does not, in a statistical sense, change significantly over a single time or space step. The Eulerian model is

$$\frac{\partial \rho}{\partial t} = D\frac{\partial^2 \rho}{\partial x^2} - \frac{\partial}{\partial x}[\rho(U) + \gamma \Phi(\rho_1, \rho_2, 2\mu)] \tag{28}$$

The variable Φ, the expected decision of the aggregate population, can be defined as

$$\Phi = \sum_{v_1=0}^{\infty} \sum_{v_2=0}^{\infty} f(v_1, v_2) g(v_1, v_2, 2\mu) \tag{29}$$

where the function f defines the joint probability density function of n_1 and n_2, and g is the social response function. At this point, we cannot go further without information about the PDF. For a suggestion about plausible assumptions about f, we turn to a plot of mean versus variance of bin density, and note that mean and variance are approximately equal in most cases (Figure 7). This suggests an assumption that the occurence of individuals is similar to a Poisson point process governed by the density

65

$r(x,t)$. Assuming that the population density at a point is governed by a Poisson distribution one can further analysis this equation. Formally, the Poisson distribution of points assumes: 1) the probability of n individuals within and interval $[x_1, x_2]$ depends only on n, x_1 and x_2, 2) the samples n_i on a distinct interval are independent, and 3) there never is an infinite number of samples on a finite interval. If these conditions are satisfied, the Poisson probability of k individuals in an interval x_1 to x_2 is given as

$$P(n=k)=e^{-\bar{\rho}}\frac{\bar{\rho}^k}{k!}, \qquad \bar{\rho}=\int_{x_1}^{x_2}\rho(x)dx \tag{30}$$

is the average number of individuals in $[x_1, x_2]$. With the Poissan point distribution, the joint probability density function for f is given as

$$f(v_1,v_2)=e^{-\bar{\rho}_1}\frac{\rho_1^{v_1}}{v_1!}e^{-\bar{\rho}_2}\frac{\rho_2^{v_2}}{v_2!}; \qquad \rho_1(x)=\int_{x-1}^{x}\rho(x)dx, \quad \rho_2(x)=\int_{x}^{x+1}\rho(x)dx \tag{31}$$

Figure 8 shows the quasi steady state numerical solution for the density of individuals at an instant (rough curve) along with mean values obtained from averaging several realization of the density distribution (smooth curve). The quasi steady-state density of individuals is maintained at a slightly higher value (47) than the optimal density prescribed in the model (40). This higher density of individuals is maintained by conditions at the sharp boundary between a large number of individuals and almost no individuals. Away from this discontinuity, the individuals do a random walk because they are unable to move effectively away from other individuals. The mean versus density variance plot (Figure 7) show that, for small mean values and variances, the data do fit the Poisson distribution (solid line in Figure 8). Figure 9 compares the numerical results of a model which uses the Poisson assumption (dashed line) and one that does not (solid line). The Figure suggests that the Poisson assumption is useful for this experiment. However, the possibility of writing PDE for the social behavior depends on the environmental properties of the behavior one is trying to simulate.

References

Alt, W. 1985. Degenerate diffusion equations with drift functions. Nonlinear Analysis, Theory, and Applications. 9:811-36.

Aoki, I. 1982. A simulation study on the schooling mechanism in fish. Bull. Japan. Soc. Sci. Fish. 48:1081-92.

Cohen, D. S. and J. D. Murray. 1981. A generalized diffusion model for growth and dispersal in a population. J. Math. Biol. 12:237-249.

Kawasaki, K. 1978. Diffusion and formation of spatial distributions. Math. Sciences 16:47-52. (Japanese).

Murray, J. D. 1993. Mathematical Biology, 2nd ed. Springer-Verlag.
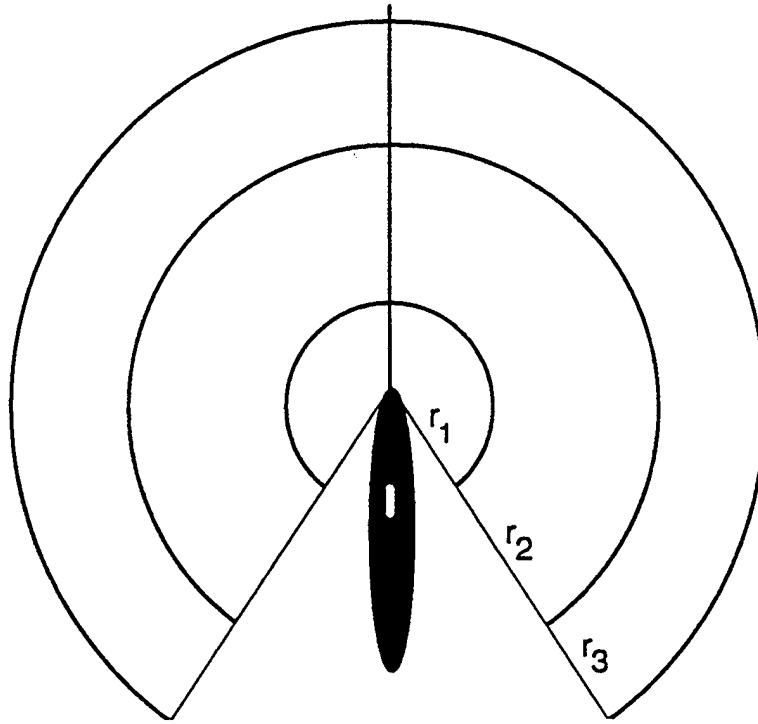
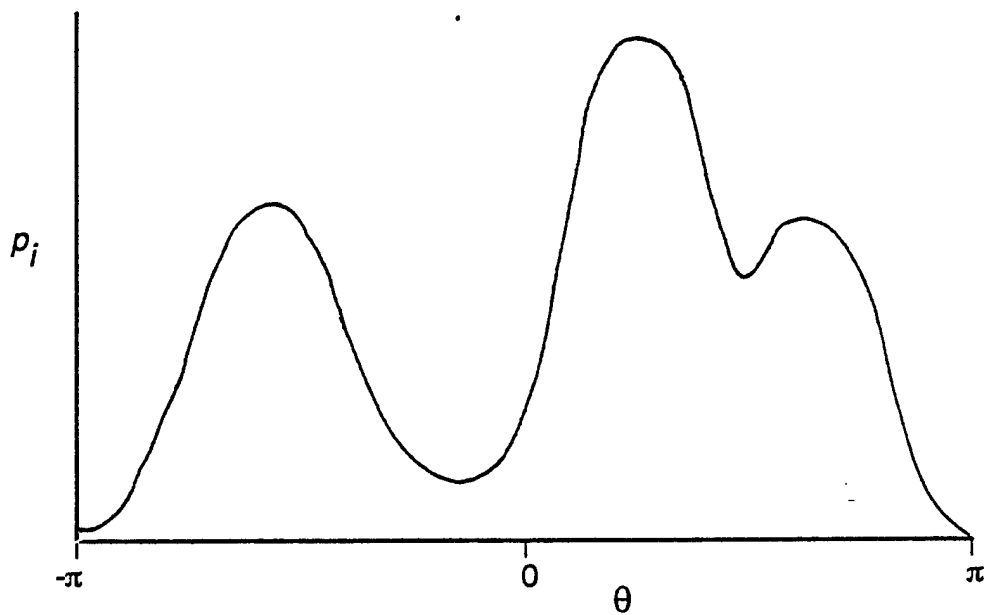Figure 1: Perception zones defining "comfort" zones around focal fish, see text for details.

Figure 2: Probability density function for the direction of movement of individual fish based on the turning angle θ. This function is a composite of   normal distributions.
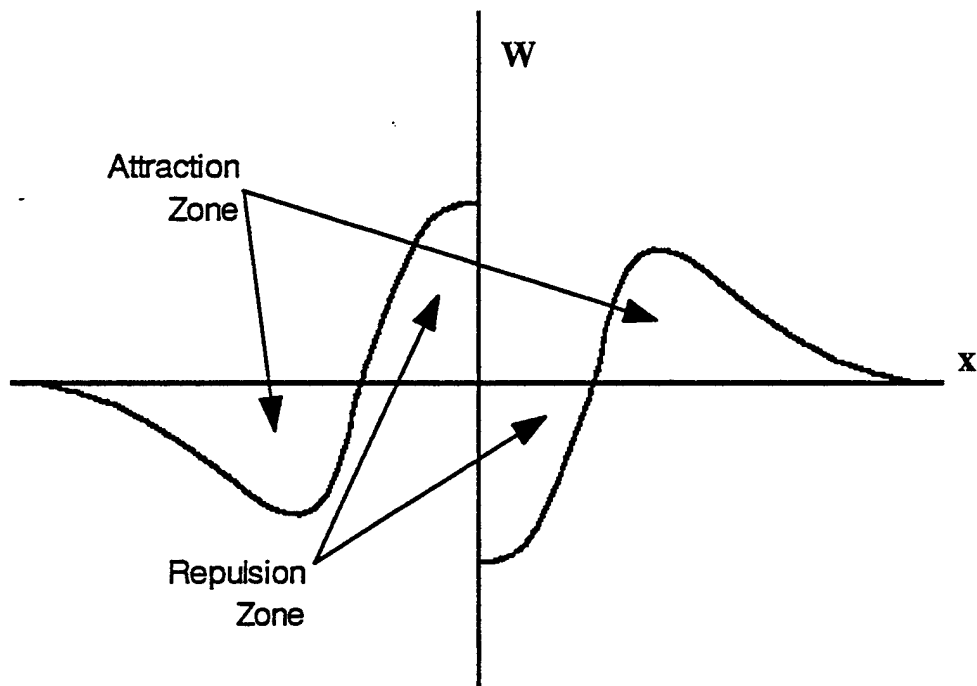
Figure 3: Response of focal individual (W) to the density neighbors as a function of distance (x).
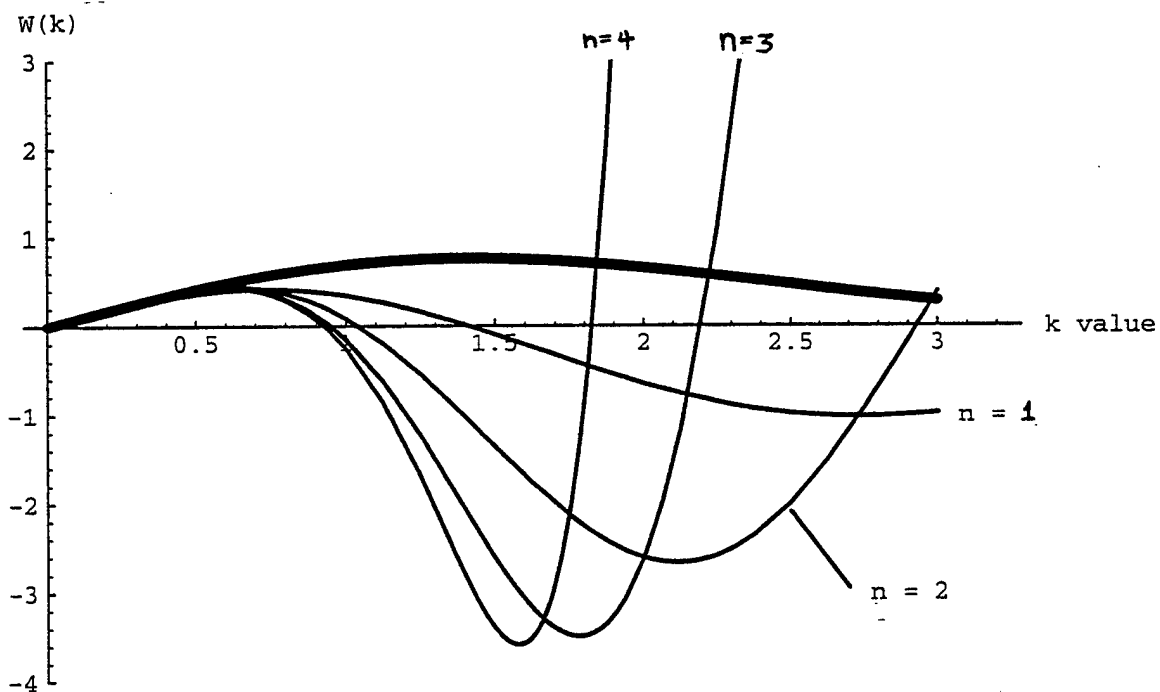

Figure 4. Curve showing the limitation in trying to simplify the integral expression to a partial differential term.
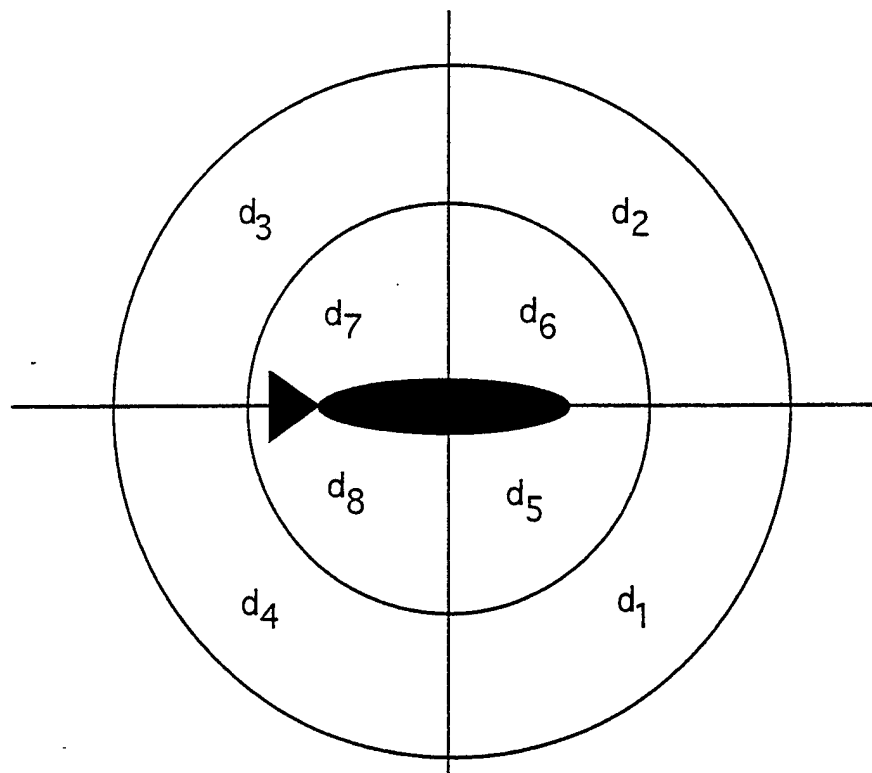
Figure 5. Fish sensing domains.
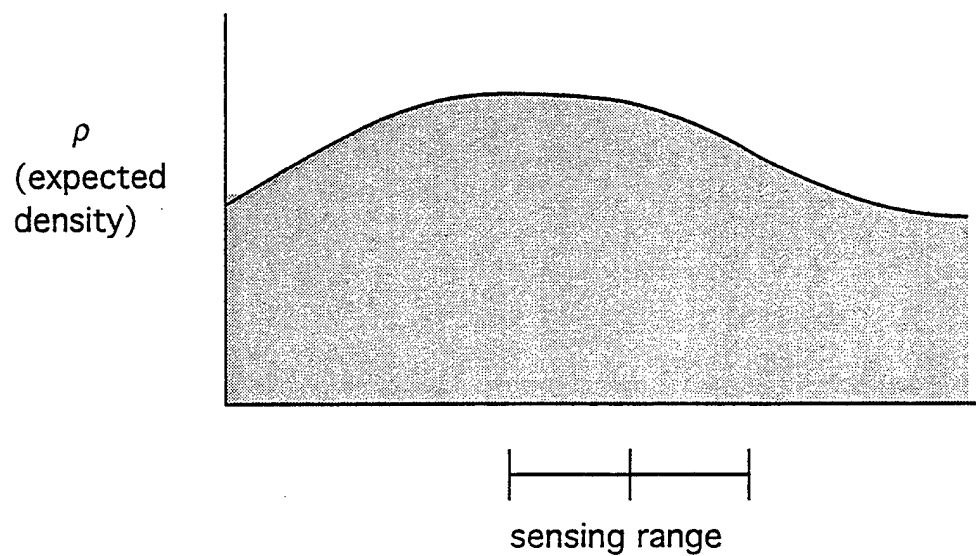
Figure 6. Density of individuals and their corresponding sensing range

Figure 7. Plot of mean versus density variance.

Figure 8. The quasi steady state numerical solution for the density of individuals at an instant (jagged line) and the mean curve obtained from averaging several realizations (smooth line).

71

Figure 9. Numerical solution of the model (solid line) and the numerical solution obtained using the Poisson distribution (dashed line)

Animal Interactions

D. Grünbaum Lecture 3
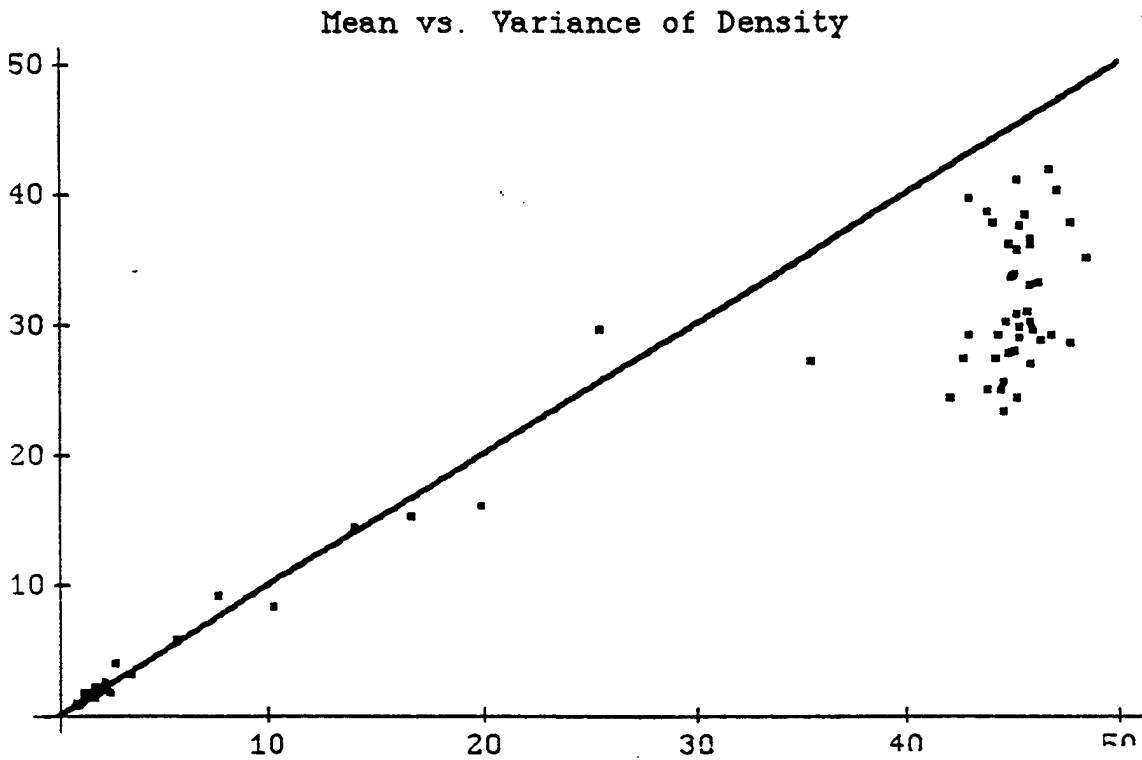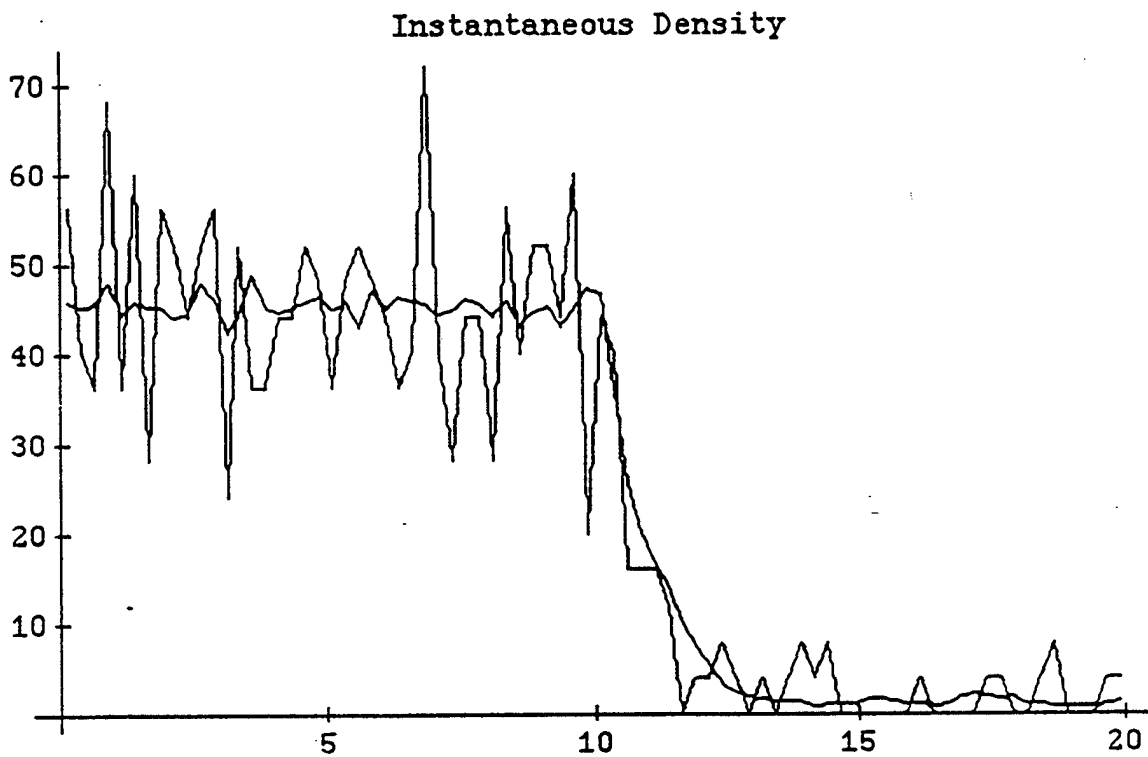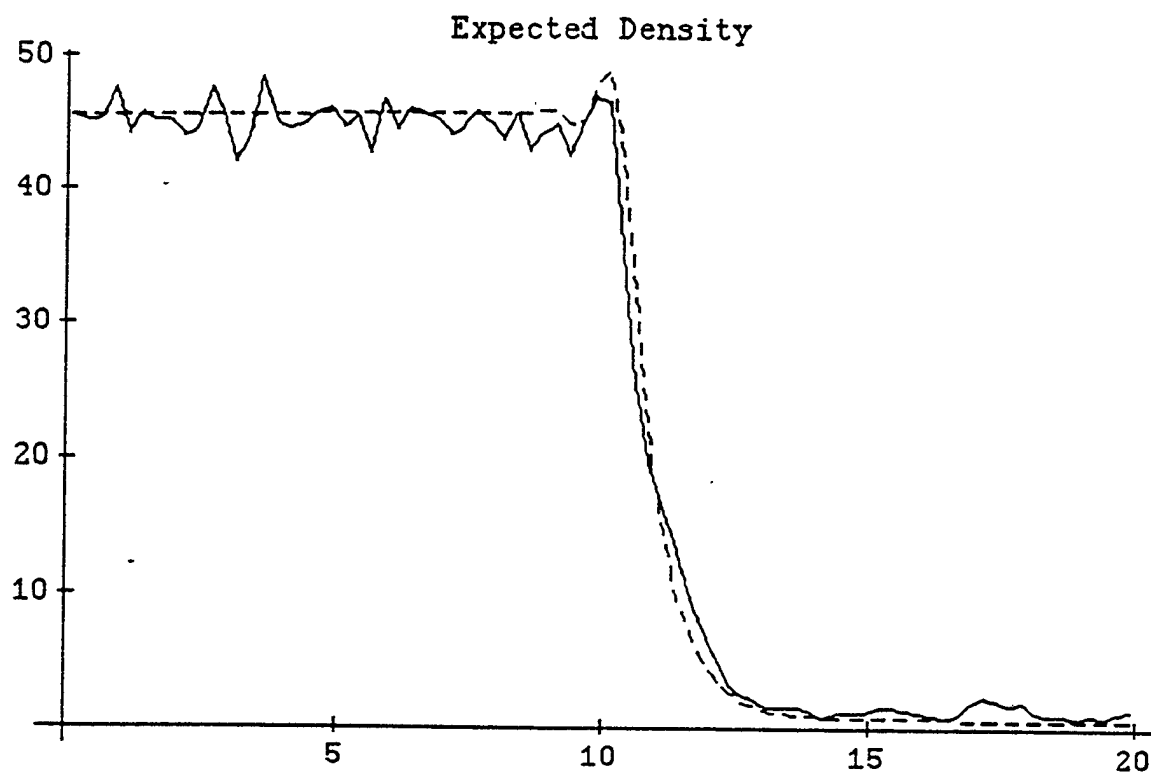
Friday

1 Non-spatial models

One of the challenges in understanding the ecological effects of social aggregations is that the dynamics of animal aggregation take place at a wide variety of length and time scales. The length scales of interest range from inter-individual distances, often on the order of centimeters, to regional differences in population density that occur on the order of hundreds of kilometers. Analogous time scales range from seconds to months. Mathematical approaches that explicitly incorporate both extremes of this spectrum are difficult to devise.

Before delving into detailed analyses of individual behavior, it is worthwhile to consider a broader view of aggregation and its biological consequences. In this broad perspective, we consider that many of the primary ecological effects of aggregation (predation and consumption rates, population movements, etc.) are determined principally by the way individuals are arrayed into groups, i.e., the group size distribution. Furthermore, there appears to be some commonality in the group size distributions of many widely differing types of animals, which suggests that details of the social behavior may not always be important (Okubo, 1986). Initially, therefore, we omit explicit consideration of the interactions between individuals, and instead focus on the interactions of groups: fusion and fission, migratory patterns, encounter rates with predators and prey, etc. In doing so, we acknowledge that ultimately we will have to justify our group-level dynamics as appropriate descriptions of the individual-level interactions. This can be done either with individual-based simulations or with some of the PDE approaches discussed elsewhere. In the meantime, this non-spatial, group-level approach gives us a context in which to understand more detailed models of aggregations, and perhaps can give us some insights into some of the common features of group size distributions across many animal taxa.

Prior to consideration of spatial distributions of animals in relation to the physical environment it is necessary to consider the basics of animal grouping, i.e. the dynamics behind aggregation. The first goal is to come up with a scheme for computing group size, n. Following the developments in Okubo (1986) the number of groups, g_n, of containing n individuals has some distribution as in Fig. 1a. The changes in this distribution of groups depends on the rate of fission and fusion of groups. These can
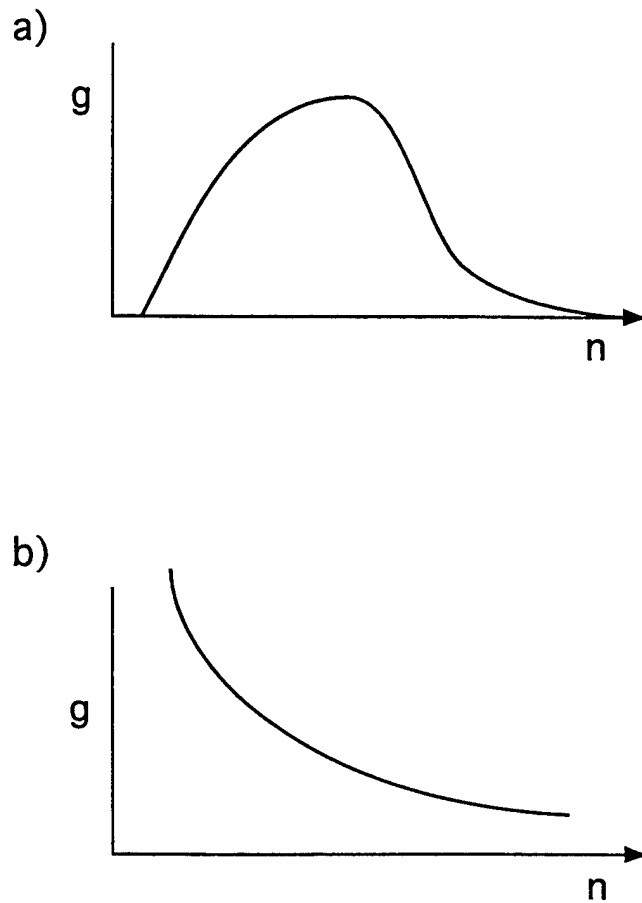
73

Fig. 1. a) Distribution of groups g in number of groups per size of group, n. The distribution drawn suggest there are no solitary individuals a broad peak and then very few large groups. b) Group distributions based on a geometric distribution. See text for details.

be written as $a(k, l, t)$, the rate of fusion of groups of size k and l into a group size $k + l$ and the rate of fission, $b(k, l, t)$ of groups of size k into two groups size l and $k - l$. The group size distribution then is governed by,

$$\frac{dg_j}{dt} = (fusion\ to\ form\ groups\ size_j)$$

$$- (fusion\ of\ groups\ size_j\ with\ others)$$

$$+ (groups\ splitting\ to\ form\ groups\ size_j)$$

$$- (groups\ size_j\ splitting)$$

In terms of the fusion and fission rates this becomes,

$$\frac{dg_j}{dt} = \frac{1}{2} \sum_{k=1}^{j-1} a(k, j-k, t) g_k g_{j-k} - \sum_{k=1}^{\infty} a(k, j, t) g_k g_j$$

$$+ \sum_{l=j+1}^{\infty} b(l, j, t) g_l - \sum_{k=1}^{j-1} b(j, k, t) g_j.$$

For a single group size this simplifies to

$$\frac{dg_1}{dt} = -g_1 \sum_{k=1}^{\infty} a(k, l, t) g_k + \sum_{k=2}^{\infty} b(k, l, t).$$

Cohen (1971) considers the case where only individuals leave or enter the groups. In this simple case

$$a(1, l, t) = a_0 + a_1 l$$

$$b(l, l-1, t) = b_0 + b_1 l$$

where a_0, a_1, b_0 and $b_1 > 0$ and $a_i, b_i = 0$ for $i = 2...\infty$

$$\frac{g_j}{dt} = [a_0 + a_1(j-1)] g_{j-1} - (a_0 + a_1 j) g_j$$

$$+ [b_0 + b_1(j+1)] g_{j+1} - (b_0 + b_1 j) g_j$$

where $j \leq 2$

75

$$\frac{g_1}{dt} = -a_0 \sum_{i=1}^{\infty} g_i - a_0 g_1 - a_1 \sum_{i=1}^{\infty} i g_i + b_0 \sum_{i=2}^{\infty} g_1 + b_0 g_2 + b_1 \sum_{i=2}^{\infty} i g_i + 2 b_1 g_2$$

Now for the case of a constant total population, N,

$$N = \sum_{j=1}^{\infty} j g_j = constant$$

the group distribution can be solved for.

As an example consider the case $b_0 = 0, a_1 < b_1$ If G = total number of groups

$$g_j = \frac{\Gamma(j + \rho - 1)}{\Gamma(\rho)\Gamma(j)}(1 - \beta)^p \beta^{j-1}$$

for $j = 1, 2, 3, ...$ with $\rho = a_0/a_1$ and $\beta = a_1/b_1$. Note that β is alway s < 1

As ρ approaches 1, g approaches a geometric distribution (Fig. 1b)

$$g_j = \left(\frac{1 - \beta}{\beta}\right)\beta^j.$$

At this point it is worthwhile considering some fundamental questions. For example, is there a maximum group size? If so, what controls its scale? On one hand one can argue that the large scale physical aspects of the environment can set the scale of groups. On the other side, however, the small scale individual behavior may dominate the fission and fusion of groups and therefore determine group size. One manner in which the latter is manifest is in density dependent behavior.

2   Encounter Rates

An example of how the fusion rates in the previous section might be made "rigorously mechanistic" (i.e., justifiable at the individual level) is an analysis of the encounter rates between groups whose spatial dimensions, velocities, directional persistences, tactic coefficients, and other characteristics vary with group size. In this analysis, we can borrow some key results from operations research that have been applied to understanding the rates of interaction for randomly moving individuals. Specifically, these results relate velocity and interaction distances to the dynamics of predator-prey systems.

Encounters between predator and prey are dependent on their relative motion whether induced by flows or by swimming. Typically the vulnerability of a prey to predation changes throughout the development of an individual. An example of this is apparent in the vulnerability of the copepod *Calanus pacificus* to different predators as shown in Fig. 2 (Greene, 1986). In the smaller napliur
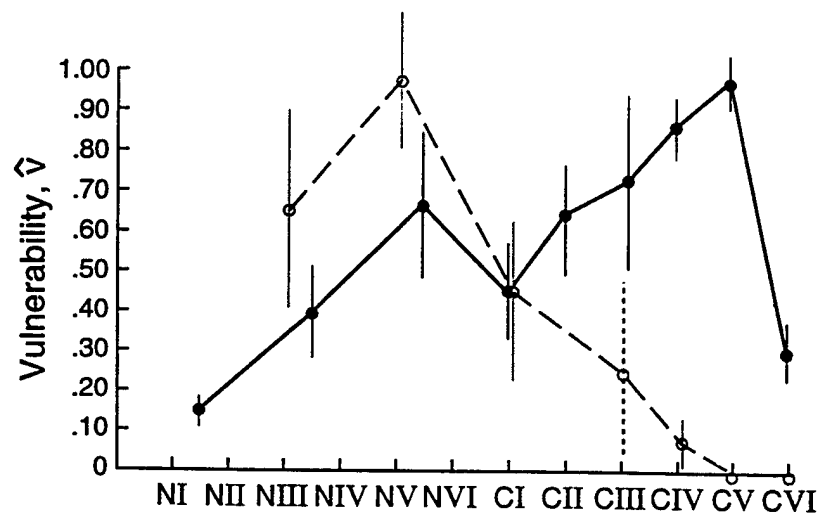
76

Fig. 2. Vulnerability to attack of *C. pacificus* by large copepods (dashed line) and ctenophores (solid line) as a function of development stage, naplii (I to VI) and then copepodites (I to VI). Figure is redrawn from figures in Greene (1986).

stages *C. pacificus* becomes more vulnerable to attack from larger copepods as it increases it size. It becomes more capable at evading attack as it develops into the copepodite stages. Vulnerability then shifts to ambush predators such as ctenophores (Greene, 1986). Switches in predator vulnerability or in prey selection over the development of a predator depends upon the relative velocities of predator and prey and the distance they can perceive each other. This distance is the encounter radius, R, and the problem becomes one of specifying the probability of encounters along with their outcomes.

Gerritsen and Strickler (1977) lay out a basic model framework for considering encounters between predator and prey. The formulation begins by considering the number of prey in some solid angle, $d\omega$ in the predators sensory field,

$$d\omega = \sin\theta d\theta d\phi$$

where the position relative to the predator is given by $\Omega = (\theta, \phi)$. The predator and prey are assumed to occupy single points in space and are randomly distributed with respect to position and direction of motion. Then the probability of a prey within the encounter radius of the predator being within the element $d\omega$ is

$$P(\Omega)d\omega = \frac{1}{4\pi}\sin\theta d\theta d\phi$$

The expected number of prey is the

$$E[n] = \frac{N_H}{4\pi}\sin\theta d\theta d\phi$$

where N_H is the density of the prey.

What remains is a specification of velocities for the predator, $\vec{v}$ and prey, $\vec{u}$. Given these their relative vector velocity is $\vec{w} = \vec{v} - \vec{u}$. The magnitude of the relative velocity is then $w = |\mathrm{w}| = \sqrt{u^2 + v^2 - 2uv\cos(\theta)}$. The rate of prey encounter is just the integral

$$Z_p = \frac{\pi R^2 N_H}{4\pi}\int_0^{2\pi}\int_0^{2\pi} w\sin\theta d\theta d\phi.$$

$$= \frac{\pi R^2 N_H}{6}[(u+v)^3 - |u-v|^3/uv].$$

This can be approximated by

$$Z_p = \frac{\pi R^2 N_H}{3}\left(\frac{u^2 + 3v^2}{v}\right); u \leq v$$

and

$$Z_p = \frac{\pi R^2 N_H}{3} \left(\frac{v^2 + 3u^2}{u} \right) ; v \leq u$$

The encounter rate and encounter probability as a function of predator and prey velocities are displayed in Fig. 3. The conclusion is that it is advantageous for a predator to either cruise at its maximum speed or to stay stationary. Hence the two behaviors seen by Greene (1986) are both reasonable strategies for a predator depending on the relative velocity of the prey. For fast prey it is best to lay in ambush, while for slow prey the encounter rate goes up with predator speed.

There are several modifications to the Gerritsen and Strickler (1977) that can be proposed. For example, the encounter radius could be made a function of the prey velocity, $R = R(u)$ to take into account the influence of a moving prey on visual or olfactory sensory capabilities. Another addition is the influence of flow in the environment on encounters as proposed by Rothschild and Osborne (1988). This theory shows that turbulence can enhance encounter rate by adding to the relative velocity. If W is the turbulent velocity the encounter rates become

$$Z_p = \frac{\pi R^2 N_H}{3} \left(\frac{u^2 + 3v^2 + 4W^2}{(v^2 + W^2)^{\frac{1}{2}}} \right) ; u \leq\leq v,$$

$$Z_p = \frac{\pi R^2 N_H}{3} \left(\frac{v^2 + 3u^2 + 4W^2}{(u^2 + W^2)^{\frac{1}{2}}} \right) ; v \leq\leq u,$$

for the two behavioral limits.

3   Density Dependent Population Process

The idea that the dynamics of large numbers of interacting animals can be described neatly by a compact set of population-level differential equations is probably the most basic justification for the study of mathematical biology. However, the success of differential equation models in describing natural biological systems has been decidedly mixed. This is in part because the ODE and PDE models have been for the most part invented heuristically as population descriptions, rather than derived rigorously from individual-based models. A rigorous derivation would not only give us more confidence that we have the best possible DE description, but presumably also would give us indications of when the approach is likely to succeed or fail. However, these derivations are difficult to carry out, especially in the case of density-dependent interactions such as predation or social interaction. Here, we examine in some detail a comparison of an individual-based simulation to various differential equation
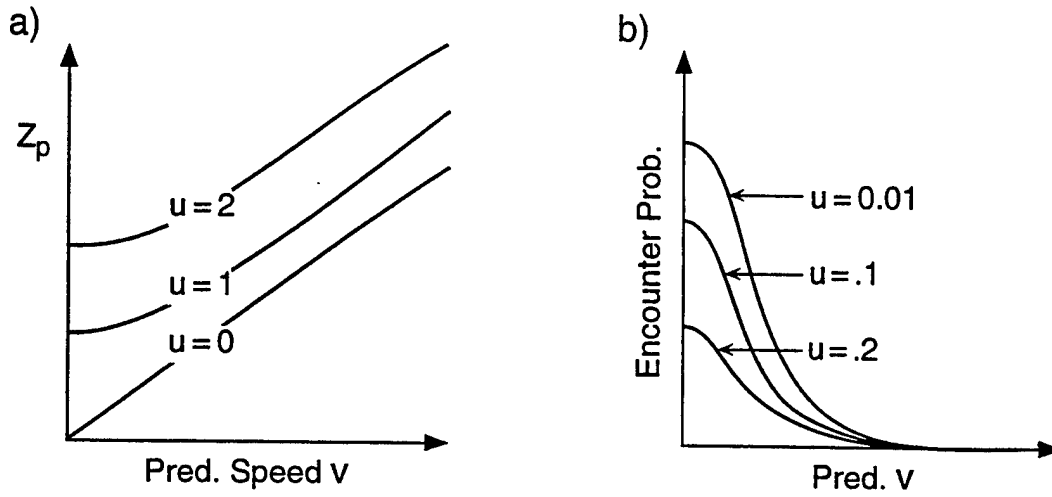
79

Fig. 3. Encounter rate, Z_p, and encounter probability as a function of predator velocity v for different prey velocities, u, in the Gerritsen and Strickler (1977) model.

approximations to it, which highlights some of the ways that essential dynamics can be compromised by the transition to differential equation descriptions.

Density dependence in population dynamics can be thought of as arising through competition between individuals for food as a predator population becomes dense or it might arise because of competition for other resources such as nesting sites. There is a large literature on density dependence and its formulation (see Murray, 1989). Here the interaction between a predator-prey system that has antagonistic predators that attack each other also is considered. The system is known as the hawk-dove problem after Durret & Levin (1993).

The hawks (H) and doves (D) problem is setup with a pay-off matrix that determines the interaction between the the populations both internally (D,D or H,H) and externally (H,D). The interactions can be written in a matrix form as

$$
\begin{array}{ccc}
 & \text{player 2} & \\
 & H \quad\quad D & \\
\text{player} \quad H & a \quad\quad b & \\
1 \quad\quad\quad D & c \quad\quad d &
\end{array}
$$

Player 1 meets player 2 and the coefficients a,b,c or d determine the pay off. For a fraction of hawks, p, then the gain for hawks is

$$\delta H = ap + (1-p)b$$

while for the doves the payoff is

$$\delta D = cp + (1-p)d.$$

For the case where hawks dominate over doves and also eliminate each other, $a < 0$ and $c < 0$ and the other coefficients are positive, i.e. hawks gain from the doves and doves are mutually beneficial. What is left to specify is the manner in which encounters occur. Here three situations are considered, 1) a dynamical system operating on the population as a whole, 2) a spatial system with Laplacian diffusion, and 3) and interacting particle system on a grid.

3.1 Version 1 : Dynamical System

The dynamical system is assumes that the interactions occur in proportion to the total population, i.e. spatial structure can be ignored. If the density of hawks is u and that of doves v, then the fraction

of hawks and doves are $u/(u+v)$ and $v/(u+v)$ respectively. The dynamics of the populations are governed by

$$\frac{du}{dt} = u\left[a\frac{u}{u+v} + b\frac{v}{u+v} - \kappa(u+v)\right]$$

$$\frac{dv}{dt} = v\left[c\frac{u}{u+v} + d\frac{v}{u+v} - \kappa(u+v)\right]$$

3.2 Version 2 : Reaction Diffusion Form

If the right-hand sides of the dynamical system above are written as $uR(u)$ and $vR(v)$ respectively and the model is placed in the spatial context by adding spatial diffusion the equations become

$$\frac{\partial}{\partial t}u = u\left[a\frac{u}{u+v} + b\frac{v}{u+v} - \kappa(u+v)\right] + K\nabla^2 u$$

$$\frac{\partial}{\partial t}v = v\left[c\frac{u}{u+v} + d\frac{v}{u+v} - \kappa(u+v)\right] + K\nabla^2 v$$

This is pictured as a continuous approximation to a random walk on the part of the populations across a two dimensional space.

3.3 Version 3 : Cellular Autonoma

Another alternative is to consider the populations as consisting of particles that move on a grid consisting of i,j points. The population of hawks at the ith and jth grid cell is then $\eta_{i,j}$. Similarly the population of doves in this grid cell is $\eta_{i,j}$. Look at a neighborhood N around (i,j)

$$\hat{\eta}_{ij} = \sum_{N_{ij}} \eta$$

$$\hat{\zeta}_{ij} = \sum_{N_{ij}} \zeta$$

The populations at site i,j then change by acting on the partition between hawks and doves in the neighborhood

$$p_{ij} = \frac{\hat{\eta}_{ij}}{\hat{\eta}_{ij} + \hat{\zeta}_{ij}}$$

acting through the payoff matrix as shown above.

82
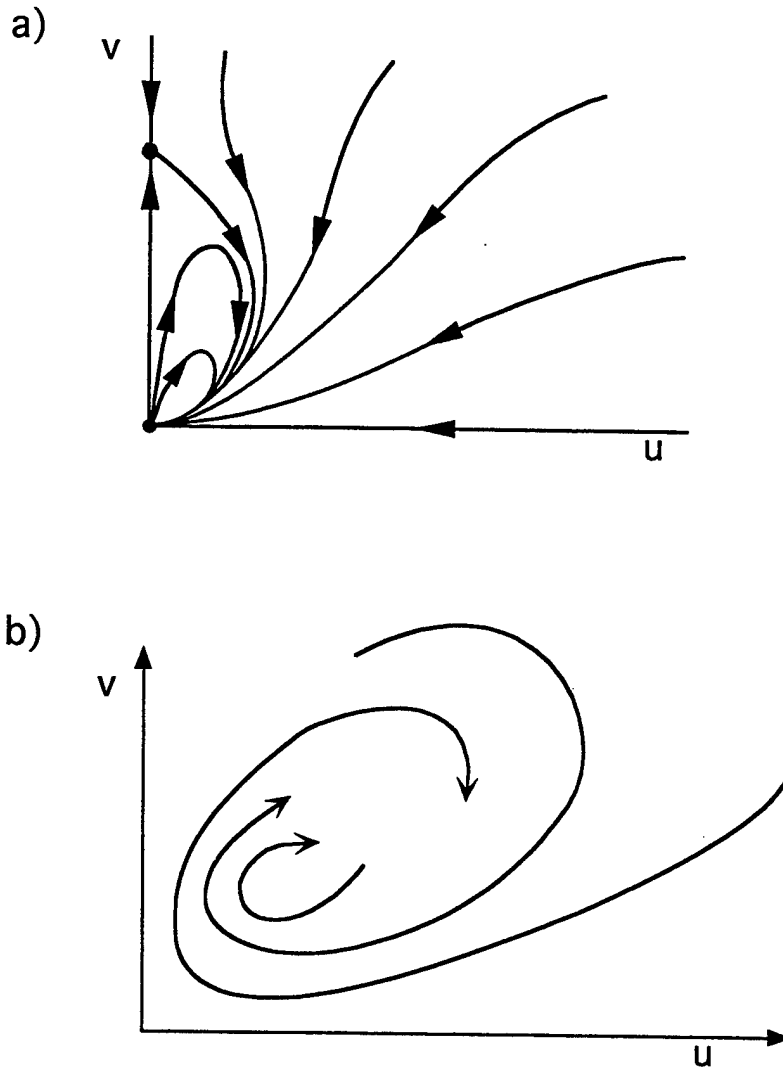
Fig. 4 a) Attractor space for the dynamic system model. The two attractors both have extinct hawks (u) and any introduction of hawks leads the attracting solution with finite numbers of doves (v) to become unstable and move to the total extinction state ($u = v = 0$). b) Attraction space for the modified dynamic system that includes explicitly probability of encounters by treating individual birds as Poisson points. See text for details.

The first two versions of the model have similar outcomes as outlined by the results from the dynamic system shown in Fig. 4a. There is a fixed state that consists only of doves. Any finite addition of hawks leads to attraction to the null state without either hawks or doves. By contrast the third version has exhibits coexistence. This is driven by persistence of dove outbreaks in the model. These do not happen persistently in the reaction diffusion case because the diffusion operator leads to infinite dispersion and has finite tails at infinity. Another way of interpreting coexistence in Version 3 is that there is a small probability of having a hawk at any given time. This allows refuge areas where doves increase. It is the finite action radius of the cell model that allows this to occur.

It is possible to modify the dynamic system to include the behavior seen in Version 3, i.e. persistence of both u and v. This is accomplished by explicitly including the probability aspect into the system. This is done by interpreting population density such that individuals represent Poisson points. The equations then become

$$\frac{\partial u}{\partial t} = u \left[a \left(h + (1 - h)\frac{u}{u+v} \right) + b(1 - h)\frac{v}{u+v} - \kappa(1 + u + v) \right]$$

$$\frac{\partial v}{\partial t} = v \left[c(1 - h)\frac{u}{u+v} + d \left(h + (1 - h)\frac{v}{u+v} \right) - \kappa(1 + u + v) \right]$$

where
$$h = \frac{1 - e^{-|N|(u+v)}}{|N|(u+v)}$$

and $|N|$ = number of grid points in N. The attractor space then takes the form sketched in Fig. 4b. This is a fair approximation of the population dynamics for the third version of the model. Adding the probability based modification is essentially taking into account the probability that a dove will encounter a hawk. This leads us to the general problem of encounter rate in an environment.

4   References

Cohen, J. E., 1971. *Casual groups of monkeys and men; Stochastic models of elemental social systems.* Harvard University Press, Cambridge, Mass.,

Durret, R. and S. Levin 1994. The importance of being discrete (and spatial), *Theor. Pop. Bio.,* **46**, 363-394.

Gerritsen, I. and J. R. Strickler, 1977. Encounter probabilities and community structure in zooplankton: A mathematical model. *J. Fish. Res. Board Can.,* **34**, 73-82.

Greene, C. H., 1986. Patterns of prey selection: Implications of predator foraging tactics., *Am. Nat.*, **128**, 824-839.

Murray, J. D., 1989. *Mathematical Biology*, Springer-Verlag, New York, 767 pp.

Okubo, A., 1986. Dynamical aspects of animal grouping: swarms, schools, flocks, and herds. *Advances in Biophysics*, **22**, 1-94.

Rothschild, B. J. and T. R. Osborn 1988. Small-scale turbulence and plankton contact rates., *J. Plankton Res.*, **10**, 465-474.

Taxis

Danny Grünbaum

Spatially-varying diffusion

We shall assume that the organisms make decisions based on the conditions at their location and/or the conditions at some distance Δx away. We shall derive continuum equations for the probability density in space and time, using the Fokker-Planck equation. If $S(x,t)$ is the probability density for the animal occupying x at time t and $\phi(x',t|x,t+\Delta t)$ is the probability that it will move from position x' at time t to position x at time $t+\Delta t$, we have

$$S(x,t+\Delta t) = \int S(x',t)\phi(x',t|x,t+\Delta t)dx' \tag{1}$$

Note that $\int \phi dx = 1$.

The organism needs to decide when to jump and which direction. We can imagine a set of criteria based on the current position or upon the ending position after the movement. If we define θ to be the relative weight of the starting vs. the ending position in the decision ($0 \leq \theta \leq 1$), we can model ϕ by

$$\begin{aligned}
\phi =&\, C_\ell([1-\theta]x' + \theta x)\delta(x' - \Delta x - x) + C_0(x')\delta(x - x') \\
&+ C_r([1-\theta]x' + \theta x)\delta(x' + \Delta x - x)
\end{aligned} \tag{2}$$

where Δx is the size of the jump and C_ℓ, C_r are the probabilities of moving left or right, respectively, and C_0 is the probability of staying put. Using this in (1) gives the evolution of the probability as a difference equation

$$S(x,t+\Delta t) = S(x+\Delta x,t)C_\ell(x+\Delta x-\theta\Delta x)+S(x,t)C_0(x)+S(x-\Delta x,t)C_r(x-\Delta x+\theta\Delta x) \tag{3}$$

In addition, we impose the constraint that the sum of probabilities of leaving a point x' to the left, the right, or staying put must sum to one:

$$C_\ell(x' - \theta\Delta x) + C_0(x') + C_r(x' + \theta\Delta x) = 1 \tag{4}$$

We use (4) to eliminate C_0 in (3) and then Taylor-expand the left-hand side to order Δt and the right-hand side to order Δx^2. Subtracting $S(x,t)$ from both sides and dividing by Δt gives the advection-diffusion equation

$$\frac{\partial}{\partial t}S = -\frac{\partial}{\partial x}(uS) + \frac{\partial^2}{\partial x^2}(DS) - 2\theta\frac{\partial}{\partial x}(S\frac{\partial}{\partial x}D) \tag{5}$$

with

$$u = [C_r(x) - C_l(x)]\frac{\Delta x}{\Delta t} \quad , \quad D = \frac{\Delta x^2}{2\Delta t}[C_r(x) + C_\ell(x)]$$

If the transitions are symmetric [in the sense that $C_r(x) = C_\ell(x)$], the steady state solutions satisfy $S = D^{2\theta-1}$ and when

$\theta = 0$	(jump determined by starting point)	$S \sim 1/D$
$\theta = 1$	(jump determined by ending point)	$S \sim D$
$\theta = 1/2$	(jump determined by half-way point)	$S \sim$ const.

In the first case, known as a "repulsive random walk" since the characteristics of the jump are determined by the conditions currently sensed, the animals accumulate in regions of low diffusivity, whereas they cluster where the diffusivity is high in the case of the "attractive random walk." The $\theta = 1/2$ case behaves like physical diffusion (or, rather, parameterizations thereof) and does not lead to any accumulation. The repulsive random walk appears to me the most common behavior.

Searching algorithm

Alt (1980) developed a model incorporating searching behavior. Let $\sigma(t, \mathbf{x}, \mathbf{u}, \tau)$ be the density of searchers at point $\mathbf{x}$ at time t who are traveling at speed $\mathbf{u}$ and have been moving at that speed for a time τ. At each time, the organism can choose to continue traveling in the same direction or to make a turn. The speed $c = |\mathbf{u}|$ is taken to be constant. This behavior corresponds to the prototypical "runs and tumbles" of simple bacterial chemotaxis. In the terminology of Othmer et al. (1988), it is a *velocity-jump process*, meaning a random walk in which individual position is continuous in time but velocity is discontinuous.

One might expect that in an effective gradient-climbing random walk, the probability of turning may be low when the animal starts out in a particular direction, but increase with time if the environment is becoming less favorable. On the other hand, it may remain low or decrease if the environment is improving. The length of run variable, τ, is introduced to allow such effects. The rate of turning is denoted $\beta(t, \mathbf{x}, \mathbf{u}, \tau)$. The arguments of the turning rate implicitly define the environment experienced by the individual since the previous turn, because they make it possible to reconstruct where the individual was at any given $\tau \geq 0$. Thus β defines the behavioral response to the variable environment, in a general form but with the limitation that an individual's "memory" only goes back as far as the last turn.

Following in the lines of the Von-Foerster equation derivation leads to the evolution equation for σ:

$$\frac{\partial}{\partial t}\sigma + \frac{\partial}{\partial \tau}\sigma + \mathbf{u} \cdot \nabla\sigma = -\beta\sigma$$

with

$$\sigma(\mathbf{u}, 0) = \int \int \beta(\mathbf{u}', \tau)\sigma(\mathbf{u}', \tau)k(\mathbf{u}', \mathbf{u})d\mathbf{u}'d\tau$$

(with the $\mathbf{x}, t$ arguments implicit). $k(\mathbf{u}', \mathbf{u})$ is the transition probability that an individual moving at $\mathbf{u}'$ before a turn is moving at $\mathbf{u}$ afterwards. k is assumed to be symmetrical, i.e., $k(\mathbf{u}', \mathbf{u}) = k(\mathbf{u}, \mathbf{u}')$. Also, no individuals are created or lost by turning, so

$$\int k(\mathbf{u}', \mathbf{u})d\mathbf{u} = 1$$

These equations are non-dimensionalized, with $\mathbf{x}$ scaled by the length L on which the environment varies and t scaled by L/c, where c again is the cruising speed. The variable τ is scaled by the characteristic length of a run, τ_0 which is taken to be small compared to the time T over which environmental variability is felt. Thus, there is a small parameter $\epsilon = \tau_0/T$ which can be used to find approximate solutions. Scaling β by the short time scale τ_0 gives the non-dimensional equations

$$\epsilon D^{\mathbf{u}} \sigma + \frac{\partial}{\partial \tau}\sigma = -\beta \sigma \tag{6}$$

$$\sigma(\mathbf{u}, 0) = \int\int \beta(\mathbf{u}', \tau)\sigma(\mathbf{u}', \tau)k(\mathbf{u}', \mathbf{u})d\mathbf{u}'d\tau \tag{7}$$

with $D^{\mathbf{u}} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla$. Define the total rate of turning by

$$
\begin{aligned}
r(t, \mathbf{x}, \mathbf{u}) &= \int_0^\infty \beta(t, \mathbf{x}, \mathbf{u}, \tau)\sigma(t, \mathbf{x}, \mathbf{u}, \tau)d\tau \\
&= \int_0^\infty \beta(t, \mathbf{x}, \mathbf{u}, \tau)\sigma(t - \epsilon\tau, \mathbf{x} - \epsilon\mathbf{u}\tau, \mathbf{u}, 0)\exp\left[-\int ds\beta(t + \epsilon(s - \tau))\right]d\tau
\end{aligned}
\tag{8}
$$

so that

$$\sigma(\mathbf{u}, 0) = \int r(\mathbf{u}')k(\mathbf{u}', \mathbf{u})d\mathbf{u}' \tag{9}$$

Finally, we shall examine the total density for all run times:

$$\overline{\sigma}(\mathbf{u}) = \int \sigma(\mathbf{u}, \tau)d\tau$$

which satisfies [from integrating (6) and using the initial condition (7), the definition (8), and (9)]

$$\epsilon D^{\mathbf{u}}\overline{\sigma} = \sigma(\mathbf{u}, 0) - \int d\tau \beta \sigma$$

$$\epsilon D^{\mathbf{u}}\overline{\sigma}(\mathbf{u}) = \int r(\mathbf{u}')k(\mathbf{u}', \mathbf{u})d\mathbf{u}' - r(\mathbf{u}) \tag{10}$$

In the repulsive case, turning rates are determined entirely at the point of departure, and we take $\beta(t, \mathbf{x}, \mathbf{u}, \tau) = \beta_0(t - \epsilon\tau, \mathbf{x} - \epsilon\mathbf{u}\tau, \mathbf{u})$ and expand in ϵ, using, for example,

$$\beta \simeq \beta_0(t, \mathbf{x}, \mathbf{u}) - \epsilon D^{\mathbf{u}}\beta_0(t, \mathbf{x}, \mathbf{u}) + \frac{1}{2}\epsilon^2 (D^{\mathbf{u}})^2 \beta_0(t, \mathbf{x}, \mathbf{u}) + \dots$$

After some manipulation, the expression for the total turning rate becomes

$$r \simeq \beta_0\overline{\sigma} - \frac{\epsilon}{\beta_0}D^{\mathbf{u}}\beta_0$$

88

and (10) gives a closed expression for $\overline{\sigma}$. Note that this expression assumes sufficiently slow spatial and temporal rates of change of the environmental variable ρ.

In the case where the turning occurs with equal probability through an angle of $-\xi_0$ to ξ around the current travel direction, we have

$$k(\mathbf{u}',\mathbf{u}) + \begin{cases} 1/2\xi_0 & \mathbf{u}\cdot\mathbf{u}' > \cos(2\xi_0) \\ 0 & \text{else} \end{cases}$$

and ξ_0 is taken to be small, we can approximate the integral in (10) and find

$$D^{\mathbf{u}}\overline{\sigma} = \frac{\partial^2}{\partial\theta^2}\mathcal{D}\overline{\sigma}$$

where θ is the angle of travel $[\mathbf{u} = (\cos\theta,\ \sin\theta)]$ and the diffusivity is

$$\mathcal{D} = \frac{\xi_0^2}{6}\left(\frac{\beta_0}{\epsilon} - \frac{1}{\beta_0}D^{\mathbf{u}}\beta_0\right)$$

In the steady state, $\overline{\sigma} = C/\mathcal{D}$.

Suppose the environmental cue $\rho(\mathbf{x})$ increases linearly in the $\hat{x}$ direction. If we define the functional behavior of β_0, we can then calculate the diffusivity, the equilibrium distribution of $\overline{\sigma}$ and the expected up-gradient velocity $\overline{v} = \int \overline{\sigma}\cos(\theta)$. Consider two cases:

1. Area-restricted search

If the turning rate depends only upon the environmental conditions at the current spot,

$$\beta_0 = 1 + \chi_0(\rho)\ ,$$

the diffusivity becomes

$$\mathcal{D} = \frac{\xi_0^2}{6\epsilon}[1 - \epsilon\chi_0'\rho'\cos\theta]$$

2. Chemotaxis

Here, the turning rate depends upon the gradient in environmental conditions.

$$\beta_0 = 1 + \chi_1(D^{\mathbf{u}}\rho)\ ,$$

and the diffusivity is

$$\mathcal{D} = \frac{\xi_0^2}{6\epsilon}[1 + \chi_1('\rho'\cos\theta)]$$

Both forms have a diffusivity including a random walk part and a part which can lead to up-gradient motion; however, chemotaxis can be much more efficient. Why do many organisms exhibit area-restricted search behavior, instead? Chemotaxis requires a way to sense gradients, whether directly or by memory of past conditions, and such capabilities may be beyond the organism. In addition, when hunting discrete prey, it may be very difficult to assess gradients or even local densities. Steps over long distances may be required, and many organisms, while capable of rapid escape movements, do not maintain speed and direction very long. There are energetic costs to exploring large spatial scales, as well. One might anticipate animals settle for a mixed strategy, perhaps also using information gleaned from neighbors.

89

Evolutionary and ecological benefits of grouping
Danny Grünbaum, June 28, 1994

1 Introduction

There are several ideas proposed to explain why animals aggregate into groups :

- *Predator saturation :* If the predator can capture only one prey at a time, then the probability of an individual prey getting hunted decreases with the size of prey group.

- *Many eyes :* Grouping into herds increases the chance of detecting and hence fleeing from danger.

- *Predator confusion :* A large enough prey group can confuse or even frighten a potential predator.

- *Foraging for resources :* If the resource is patchy but large enough to feed many individuals when found then grouping may be a very efficient strategy for foraging.

In this lecture, we examine three examples of groupings found in nature :

- *Swarming locusts.* Locusts form huge fairly coherent swarms consisting of billions of individuals. Danny Grünbaum has constructed an individual based model to mimic the behavior of the swarm. The description and results of his model are described here.

- *Schooling fish.* In order to provide a theoretical basis for the many eyes and better foraging arguments mentioned above, we discuss the work of Grünbaum (199x). These notes make liberal use of text and figures from the source which carefully compares asocial and social behavior in a controlled experiment to bring out the advantages of grouping.

- *Aggregating slime mold.* Slime mold amoebae exist in unicellular form when there is enough food. Under stress, the cells communicate with each other using a chemical messenger, cAMP and aggregate to form fruiting bodies. Most of the cells form the stalk of the fruiting body and hence die. A few lucky ones become spores which are dispersed and hence form the basis of the next generation. In absence of this aggregation, chances are that all the unicellular organisms would die. In this way, evolution has selected for aggregation so that a few individuals survive at the expense of most when the alternative is death for everyone. The material in these notes is shamelessly lifted from Keller & Segel (1971) which discusses the conditions necessary for the instability responsible for aggregation.

2   Swarming locusts

African locusts exhibit two distinct social behaviors viz., solitary and gregarious. These are determined by population density i.e., whether each locust exhibits one behavior or the other depends on how many members of its own species it finds near itself. In the first behavior, which occurs during low density situations, the individual lives an independent lifestyle. It largely ignores its neighbors and competes for resources on its own terms.

When the locust density exceeds some critical density, however, the locusts' behavior changes dramatically. The locusts begin to respond to olfactory cues which serve to engage all individuals in an ensemble (or gregarious) activity. The locusts then tend to aggregate readily. At this stage, they are usually juveniles and cannot fly. They form "hopper bands" and roam across the land in groups ranging in size from tens of thousands to millions of individuals. While gregarious, the individuals become adults and acquire the ability to fly. From a hopper band, the group takes the form of a rolling swarm. Both hopper bands and swarms are very destructive to the flora of land they pass over. This can have serious consequences for agricultural areas they pass over.

Examination of the behavior of a single locust initially on the ground yields insight into the dynamics of the swarm. The locust first flies into the wind as it takes off from the trailing edge of the swarm. Then, while near the top of the swarm, it is carried downwind at a typical airspeed of around 15 m/s. After some amount of time in air, it descends, heading back into the wind as it drops to the leading edge of the swarm. When many locusts each behave in this way, the swarm as a whole adopts an inverted trapezoidal shape, short side down. Each locust spends on the order of a half hour on the ground, resting and foraging for food before taking off again. Because of the merging of several ground groups, the locust swarms can consist of 10^9 to 10^{10} individuals and have a footprint of 10s to 100s of square miles. The density of the swarm is roughly one locust per cubic meter, plus or minus an order of magnitude. As a group, they move downwind at a rate of a few kilometers per hour. The form of the swarm is highly variable, the above being a description of the apparent general tendencies.

It is hypothesized that the reproductive advantage of this social strategy comes from finding rain since wet ground is necessary for locust reproduction. The surface winds are likely to be blowing in the direction of a surface convergence zone, above which moisture in upwardly transported air may fall as rain.

The tendency of an individual locust to head into the center from the edges is low when heading into the wind, but higher when traveling with the wind. Can this large scale behavior be explained by certain rules which each locust follows individually ? It is argued that while flying, the locusts prefer to sense a certain apparent (retinal) velocity of the ground. While close to the ground, they tend to head into the wind in an effort to reduce the rate at which the ground passes through their field of view. As they gain altitude, this apparent rate will decrease due to perspective. Therefore, the higher locust will tend to move with the wind in

order to maintain the preferred retinal velocity. Since locusts appear to have a visual acuity of about 0.5-1.0 degrees and a focus at infinity, it is reasonable to hypothesize this type of behavior.

A model has been constructed in an attempt to capture this type of large scale behavior from local decisions made by individual locusts. For this model, it is hypothesized that the locusts try to maintain a certain orientation depending on their altitude. This is modeled as a 'comfort' function. Diffusion of the locust density value is a constant plus a value inversely proportional to the comfort level. This comfort function results from both the desire for a certain apparent velocity at a given altitude and a desire to fly within a certain altitude range. Hence, the comfort function depends on flight orientation angle (relative to horizontal-downwind) and altitude. If we look at the time evolution of density of locust population versus orientation and altitude starting from a situation of all locusts on the ground and evenly distributed about orientation angle, the population density plot tends to look like the comfort function. At the altitude where the density-weighted orientation angle is equal to $\frac{\pi}{2}$, the net locust population travels at the speed of the wind. Higher, the locust flies with the wind; lower, it flies against the wind. Summing across altitude yields density versus altitude. Doing this shows that the maximum density occurs high in the swarm. In fact, more locusts are above the altitude where the locusts' average speed equals the wind speed, so the swarm as a mass moves downwind. The net result is a modeled population whose bulk properties, velocity and density with height, are in general agreement with those observed in real swarms.

3 Schooling as a strategy for chemotaxis in a noisy environment

Most aquatic animals employ tactic behaviors to follow environmental gradients and locate patchily distributed resources. Important resources, particularly those which are advected within the water column, often display ecologically significant variations at large length scales ($\geq 10km$). However, the gradients which might lead an animal to large-scale resource concentrations are not only weak but overlain by "noisy" fluctuations at the micro-scales ($\leq 10m$) at which animals typically sense their environments. Here, a theoretical basis is provided for the hypothesis that schooling behavior improves the ability of school members to climb chemical gradients under challenging conditions when directional signal-to-noise ratios are low.

First, we examine a simple, asocial bacterial chemotaxis algorithm. The results of this null model are compared with simulations in which individuals display schooling behavior in addition to chemotaxis. Effects of group size on the success of finding resources is examined. Also addressed is the responsiveness of the school to time-varying chemical gradient direction.

3.1 Asocial searching

An animal that cannot directly sense favorable regions must locate them through some sort of trial-and-error process of probing in various directions and modifying course depending on the results. Here, we present a simple Markovian model for chemotaxis. Suppose, for example, that individuals are searching on a plane surface where the concentration increases in the positive x-direction. Further, suppose that the individuals are moving at unit speed, and change their direction randomly to the left or right at discrete time intervals Δt and the difference in angular orientation is small if the individual is oriented up-gradient and larger if orientation is down-gradient i.e.,

$$\Delta \theta = r_1 \Delta \theta_1 \left(\frac{1 - \cos(\theta)}{2} \right) \tag{1}$$

where $\Delta \theta$ is the angular change in a time interval Δt, and θ is the angle with respect to the concentration gradient (the x-axis, here). r_1 is randomly chosen to be $+1$ or -1 at each time step so that an individual turns with equal probability to the left or right. $\Delta \theta_1$ is the maximum difference in angular heading the individual will make if it is headed down-gradient. Following these rules, the individual will move away from unfavorable headings much faster than from favorable headings and will soon point in the correct heading (see figure 1). If there is small-scale noise, then the animal might feel the concentration is increasing when actually it is decreasing and vice-versa. To model this effect, we modify the above rule by adding a random angular change regardless of the animal's orientation i.e.,

$$\Delta \theta = r_0 \Delta \theta_0 + r_1 \Delta \theta_1 \left(\frac{1 - \cos(\theta)}{2} \right) \tag{2}$$

where r_0, like r_1 is also chosen to be $+1$ or -1 at each time step. $\Delta \theta_0$ measures the vigor of random motion the individuals go through in absence of any signal, or equivalently the noisiness of the signal.

This type of chemotaxis has been analyzed thoroughly (Alt 1980, Okubo 1980, 1986, Othmer et. al. 1988) making it a convenient searching behavior with which to compare searching strategies of social individuals. The long-term behavior of these individuals can be expressed by taking the diffusion limit of the random walk in θ. The present case is a "repulsive" random walk, in which the probability of moving depends only on conditions at the point of departure. The probability density of orientation angles, $\rho(\theta, t)$, evolves according to

$$\frac{\partial \rho(\theta, t)}{\partial t} = \frac{\partial^2}{\partial \theta^2} D(\theta) \rho(\theta, t) \tag{3}$$

where $D(\theta)$ is a directionally varying diffusivity. $\rho(\theta, t)$ represents the probability at time t that an individual is heading in the direction θ or, equivalently, the fraction of a large (non-interacting) population with that orientation. The diffusivity has a constant and a
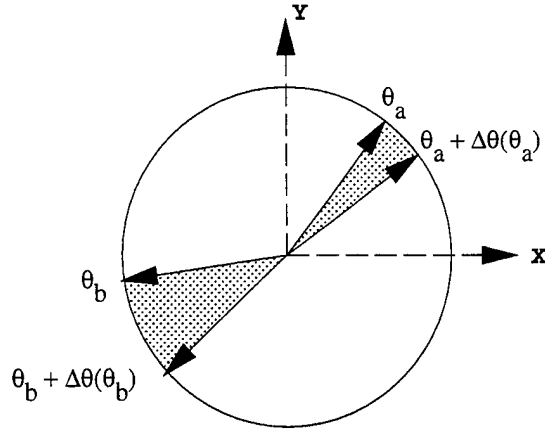
93

Figure 1: "Bacterial" chemotaxis by directionally variable "step" size in the change of heading angle. Shown are two heading direction, a nearly up-gradient orientation (θ_a) and a nearly down-gradient orientation (θ_b), where the concentration of an attracting substance increases in the positive x-direction. An individual changes heading in either direction, right or left, with equal probability. However, increments in angular orientation are smaller if the individual is oriented up-gradient and larger if orientation is down-gradient. Thus, on average, an individual spends more time moving in up-gradient than down-gradient directions, resulting in up-gradient taxis.

directionally varying component,

$$D(\theta) = D_0 + \left(\frac{1 - \cos(\theta)}{2}\right) D_1, \tag{4}$$

where D_0 and D_1 are given (Levin, 1986) by

$$D_0 = \frac{\Delta\theta_0^2}{2\Delta t}, \quad D_1 = \frac{\Delta\theta_1^2}{2\Delta t} \tag{5}$$

The steady state solution to the Fokker-Planck equation (3) is

$$\rho(\theta) = \frac{(1 + d)^{\frac{1}{2}}}{\pi(2 + d(1 - \cos(\theta)))}, \tag{6}$$

where $d = D_1/D_0$ is the relative diffusivity (Davis et. al., 1991), a measure of the relative strength of the chemotactic signal and noise — small d means that the gradient is heavily obscured by small-scale noise; large d means that the gradient is distinct. This equilibrium distribution of orientation angles shows the fraction of time a typical individual spends in each orientation, and consequently, how fast it moves up-gradient on the average. The average up-gradient velocity of an individual, as a fraction of individual's forward speed is

$$U(d) = \frac{(1 + d)^{\frac{1}{2}} - 1}{(1 + d)^{\frac{1}{2}} + 1}. \tag{7}$$

Equations (6) and (7) summarize the effectiveness of the bacterial chemotaxis algorithm — when the signal dominates noise ($d \gg 1$), individuals virtually always orient correctly, and progress up-gradient at close to full speed; when the noise predominates ($d \ll 1$), the angular distribution of individuals is nearly uniform, and the up-gradient velocity is nearly zero (see figure 2). In an intermediate range of values ($0.3 \leq d \leq 3$), there is measurable but slow movement up-gradient. Following, we address whether individuals operating in this intermediate signal-to-noise range improve their ability to move up-gradient by adopting a social behavior i.e., schooling.
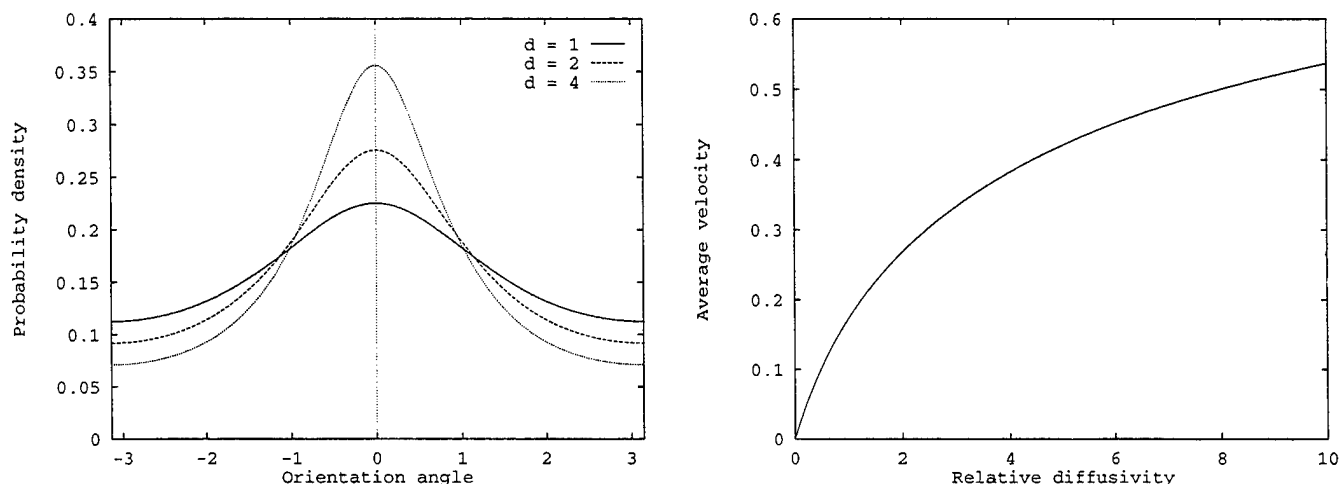
Figure 2: The effectiveness of the simple "bacterial" chemotaxis algorithm as a function of d, the relative diffusivity. The left panel shows that the equilibrium distribution of orientation angles is more tightly clustered about the up-gradient direction for higher d. The right panel shows that individuals spend a greater fraction of their time in favorable orientations, resulting in higher average velocity with increasing d.

3.2 Simulations of searching with schooling behavior

To investigate the gradient-climbing ability of schooling searchers, we employ a simplified model of schooling behavior (figure 3). Subtleties of schooling models are ignored for the sake of keeping the simulations and analysis simple. The key attributes of schooling models are :

- *Limited sensory area* : The animals have a decreasing probability of detection or responsiveness to neighbors at large separation distances. This effect is modeled by individuals responding only to other individuals within a unit radius (the "detectable" neighbors); individuals more than a unit radius are ignored. Also, equal weight is given to the detectable neighbors.

- *Density preference/Grouping tendency :* The schooling individuals have a preference for a certain density range — they will turn towards their neighbors if the school is sparse and turn away from their neighbors if the school gets too dense. This switch from attractive to repulsive interactions mediated by local density of animals is modeled as follows. If the number of detectable neighbors is within the preferred density range, then the individual takes no action. On the other hand, if the number of neighbors is outside that range, the individual turns by a small amount, $\Delta\theta_3$, to the left or right according to which side has more neighbors and whether it has too many or too few neighbors in order to bring the number of neighbors within the preferred density range.

- *Alignment tendency :* This is the tendency of animals to match velocities. Since the animals are moving with unit speed, this effect is modeled by randomly choosing one of the detectable neighbors and turning a small amount, $\Delta\theta_4$, towards that neighbor's heading. Over many time steps, the individual will tend to align with the average heading of its nearby neighbors.

- *Individual behavior :* The animals are responding to other cues in the environment such as predator, prey or resource gradients. This, in our case, simply amounts to the individuals retaining the bacterial chemotactic behavior.
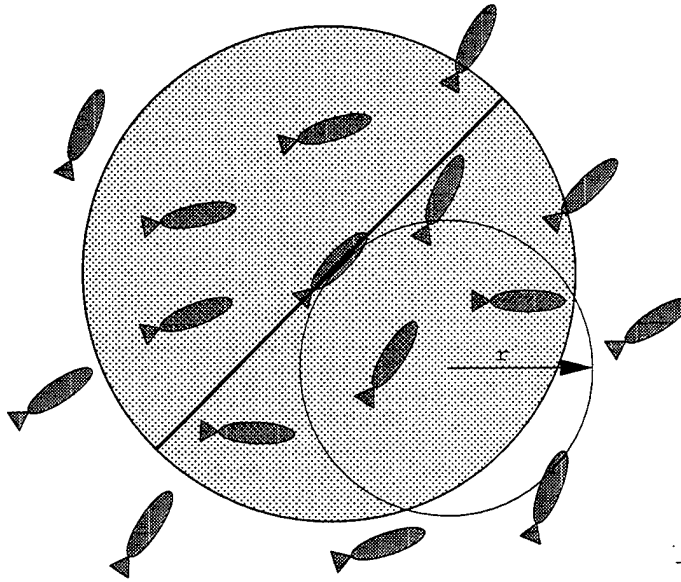

Figure 3: Schematic of schooling behavior. In the simplified schooling algorithm, each individual reacts to neighbors within a unit distance, shown here as a shaded circle. Also shown is r, the radius of the tightest turn that can result from social behavior. If this radius is smaller than the reaction distance ($r \ll 1$) then individuals can turn tightly to remain close to neighbors; if it exceeds the reaction distance ($r \gg 1$), individuals are not capable of turns tight enough to stay near a neighbor.

These behaviors can be made independent of step size by expressing them as $\alpha = \Delta\theta_3/\Delta t$, the rate of turning towards and away from neighbors, and $\beta = \Delta\theta_4/\Delta t$, the rate of turning to align with neighbors. Incidentally, the parameters α and β can also be usefully thought of in terms of

$$r = \frac{1}{\alpha + \beta} \tag{8}$$

the radius of the tightest circle that an individual can make under the combined influence of the two social behaviors. The size of this radius, relative to other characteristic lengths such as the detection range, determines in part the group properties of the school.

The results of the simulations show that schooling individuals, on average, move more directly in an up-gradient direction than asocial searchers with the same chemotactic parameters.

Figure 4 shows the individual positions in a periodic domain in simulations of asocial and social chemotaxis at $t = 100, N = 64$. The chemotaxis parameters, $D_0 = 0.3$ and $D_1 = 0.6$ are within the range of relative diffusivity ($d = 2$) for which up-gradient motion due to bacterial chemotaxis is relatively inefficient ($U = 0.268$). The social parameters are $\alpha = 6.0, \beta = 6.0$. Individuals seek to have between 56 and 63 detectable neighbors. Figure 5 displays the distribution of orientation angles at $t = 100$ in the simulations shown in figure 4, showing a tightened angular distribution in the social chemotaxis simulation compared to the asocial case. Figure 5 also shows the time-series of average up-gradient velocities for the same simulations, beginning with random initial conditions and continuing until $t = 100$. Velocity is stochastic in both cases, but higher for the social algorithm.
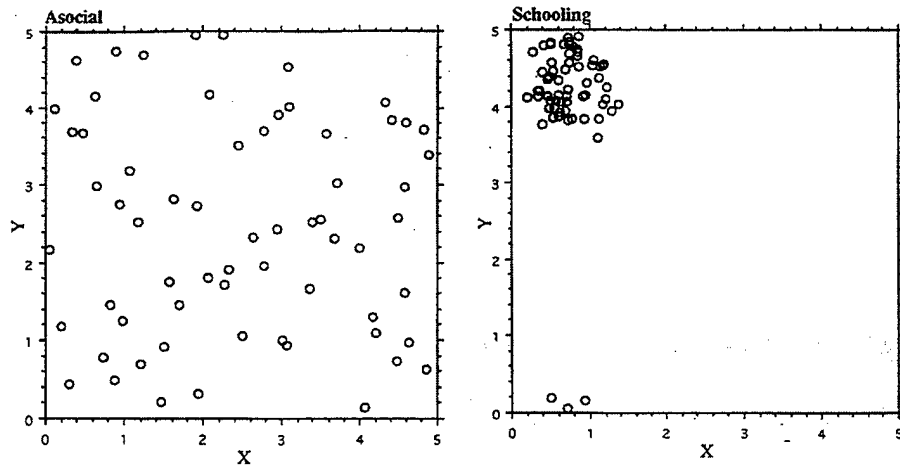

Figure 4: Simulations of asocial and social chemotaxis at $t = 100$ with $N = 64$. The left panel displays the positions of individuals for the asocial, "bacterial" chemotaxis algorithm whereas the right panel displays positions of individuals for the "social" chemotaxis algorithm. The gradient is uniform and increases to the right.
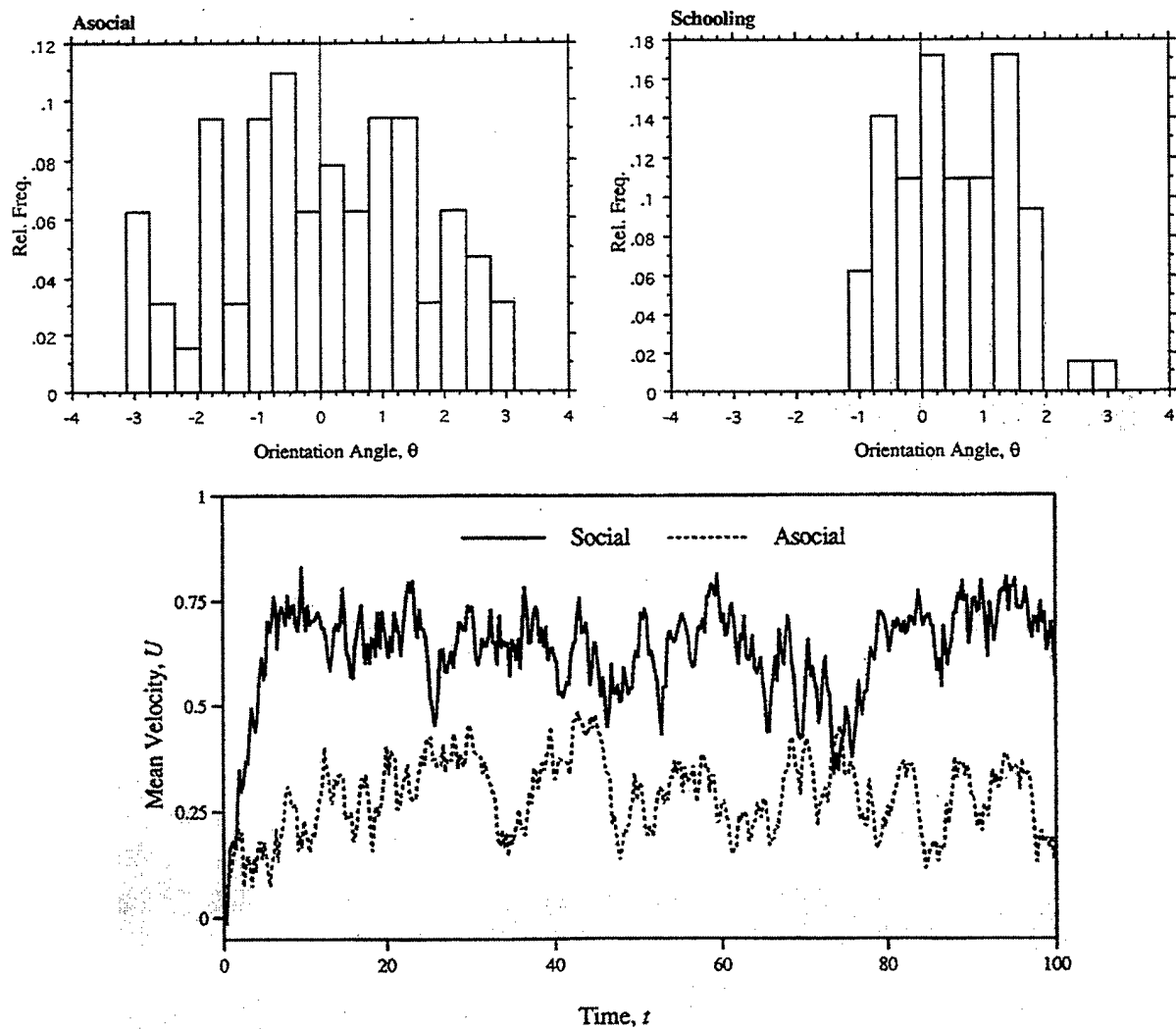
Figure 5: Simulations of asocial and social chemotaxis. Orientation angles show a tightened angular distribution in the social chemotaxis (top right) compared to the social case (top left). Bottom panel displays time series of average up-gradient velocities for the same simulations. Velocity is stochastic in both cases, but higher for the social algorithm.

98

Effect of group size

Here, we address the question of how large a group must be before the individuals start experiencing the benefits of schooling and whether there are diminishing returns at large group sizes.

Figures 6 and 7 show the gradient-climbing velocity as a function of group size. Figure 6 shows the histograms of average velocity over the interval $0 \leq t \leq 30$, for asocial searchers and social searchers in groups of 8, 16 and 32. Asocial search is represented by 1024 individuals; social searches are represented by the same number of individuals arranged in groups of several sizes (128 runs of 8 individuals, 64 runs of 16 individuals, 32 runs of 32 individuals). Figure 7 shows the median, 10th and 90th percentiles of group velocity as function of group size. Included in this plot are groups of 64 and 128 (32 runs for each). Parameters are as in figure 4, except that in each social simulation the number an individual seeks is 7/8 of the total number of individuals (e.g., 28 neighbors for $N = 32$).

For the current simulations, the group need not consist of very many individuals to provide its members with a searching advantage. Schools with as few as eight members do substantially better on average than isolated individuals. Also, we see that chemotactic efficiency increases with group size until roughly 64 individuals and that further increases in size have little effect on up-gradient velocity.

Another effect of increasing group size is to decrease the variation in searching success between individuals as seen by the difference between the 10th and 90th percentile searchers falling with increasing group size.
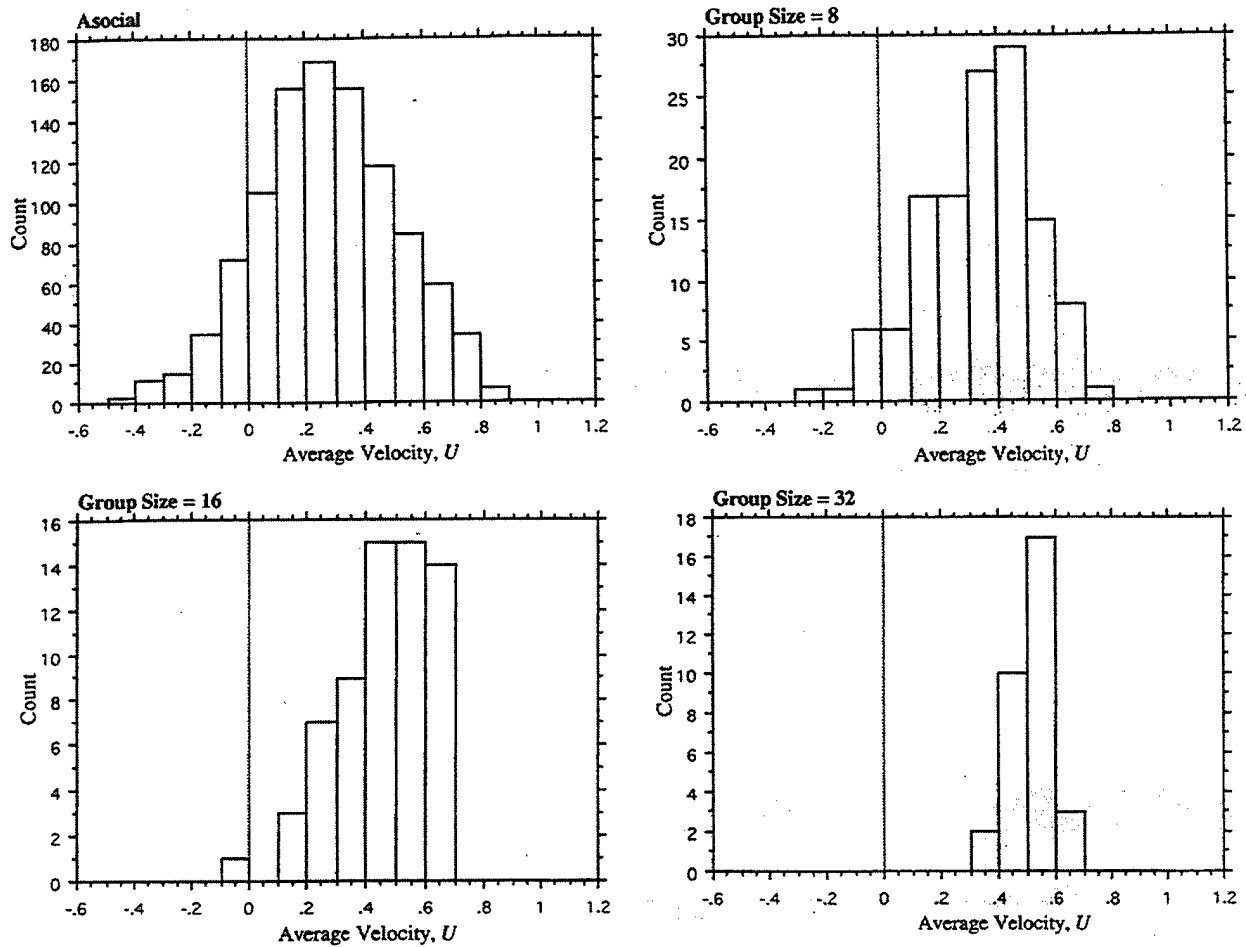
Figure 6: Gradient-climbing velocity as a function of group size. Displayed are histograms of average velocity for asocial searchers and social searchers in groups of 8, 16 and 32.
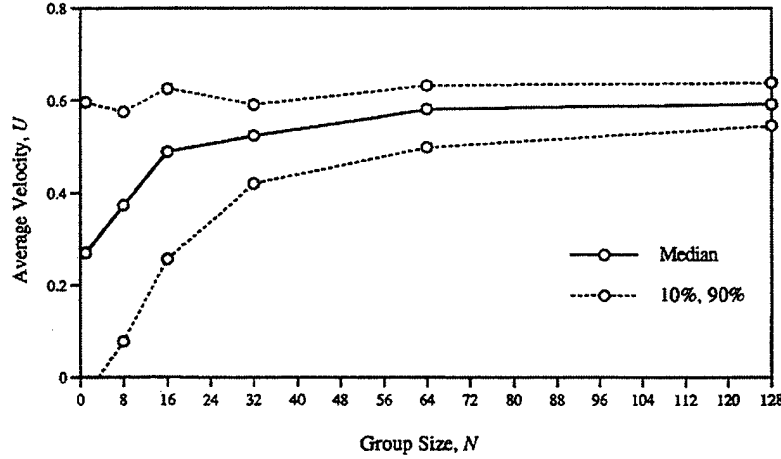
Figure 7: Median, 10th and 90th percentiles of group velocity as a function of group size.

3.3 Perfect vision : a non-spatial approximation to social chemotaxis

As the group size and density grow, stochasticity in the behavior of each member has a relatively weaker effect on group motion and the behavior of the group as a whole becomes more consistent and predictable. Consider now a school of large size and high density, and assume that the school is strongly cohesive i.e., the group always remains together and the grouping tendency is invoked relatively seldom by a typical member and can be neglected. In this dense and cohesive school, individuals might be said to have "perfect vision", because they almost "see" amongst their neighbors a large and representative sample of the angular distributions of the group as a whole. The distribution of orientation angles can be seen to evolve as

$$\frac{\partial \rho(\theta,t)}{\partial t} = \frac{\partial^2}{\partial \theta^2}D(\theta)\rho(\theta,t) - \beta\frac{\partial}{\partial \theta}\left[\rho\left(\int_0^{\theta+\pi}\rho(\theta')d\theta' - \int_{\theta-\pi}^0\rho(\theta')d\theta'\right)\right]. \qquad (9)$$

where the first term corresponds to the chemotactic behavior and the second term corresponds to the alignment behavior — an individual's equal probability of choosing any one of its neighbors at a given time. β is the rate of turning towards neighbor's headings. This non-linear partial integro-differential equation cannot be solved analytically in general and is hence solved numerically. Solutions of (9) are consistent with simulations discussed above — increasing rate of alignment, β results in more rapid convergence about the up-gradient direction and tighter equilibrium angular distributions (see figures 8 and 9).
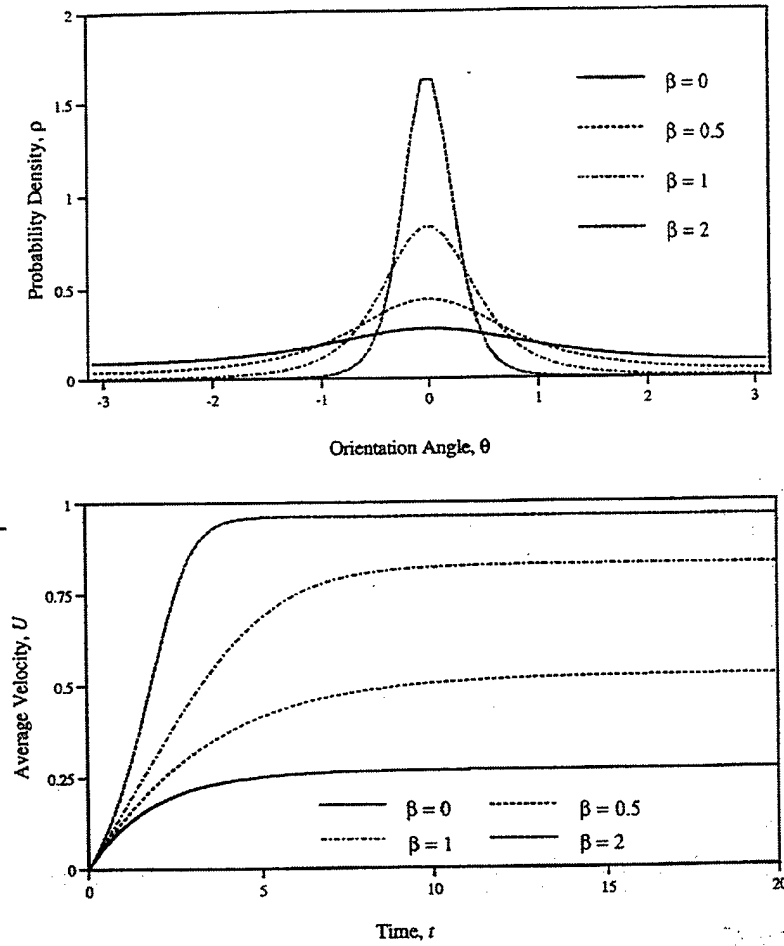
101

Figure 8: Orientation angle distribution under the "perfect vision" scenario for asocial searching ($\beta = 0$) and three choices of the alignment parameter, β (in all cases, $d = 2$). The top panel displays the equilibrium distribution of orientation angles, showing tightened distributions for stronger alignment tendency. The bottom panel shows average up-gradient velocity as a function of time with initially uniform angular distributions. Stronger alignment tendency results in a faster response to the gradient and higher up-gradient velocity.
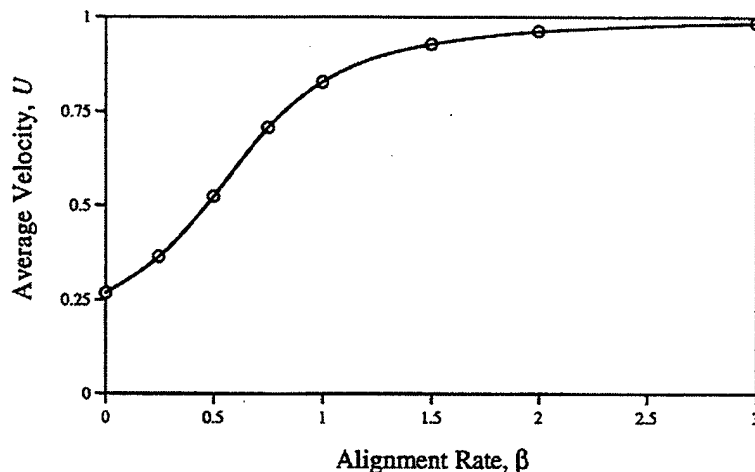
Figure 9: Average up-gradient velocity under the "perfect vision" scenario as a function of the rate of alignment, β. Velocity initially increases rapidly with β, then saturates and approaches the maximal value asymptotically.

Response to time-varying signal

The calculations so far have all considered an initially uniform angular distribution and constant (albeit noisy) chemotactic signal. In a more realistic environment, where concentration of the chemotactic substance varies in time and space, the direction of the large-scale gradient might change as the searchers swim along. To climb gradients effectively under these circumstances, searchers need the ability to adjust and correct heading direction.

In the example in figure 10, the gradient is initially pointed in the negative y-direction. After an initial period of 5 time units, during which the school orients perpendicularly to the x-axis, the gradient reverts to the usual x-direction orientation. The school must then adjust to its new surroundings by shifting to climb the new gradient. This example shows that alignment works against course adjustment — the stronger the tendency to align, the slower is the group's reorientation to the new gradient direction.

3.4 Conclusions

Social chemotaxis is essentially bound up in the statistics of pooling the outcome of many unreliable decisions — if averaging decisions among a large group of individuals at one instant can substitute in part for averaging many decisions by a single individual over space and time, then each member of the group may arrive at the "right" decision more quickly and with greater accuracy than it would in isolation.
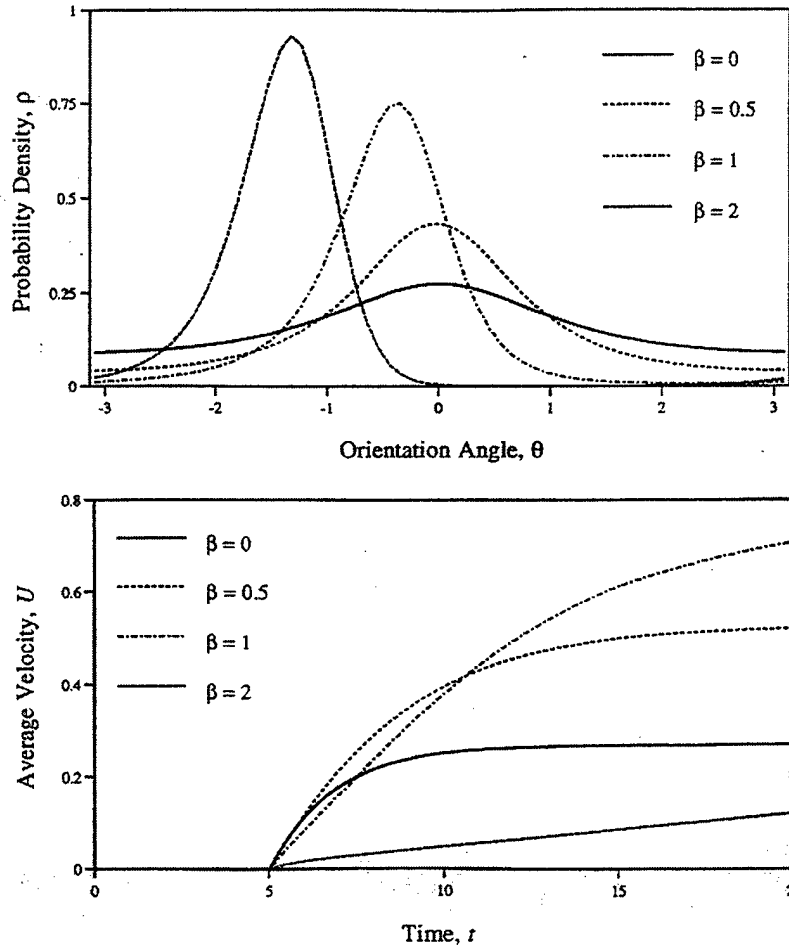
Figure 10: Response to changing gradient directions under the "perfect vision" scenario for asocial searching ($\beta = 0$) and three choices of the alignment parameter, β. These calculations are the same as those in figure 8, except that the gradient is along the y-axis ($\theta = -\pi/2$) for the time interval $0 \leq t \leq 5$, and reverts to the usual orientation ($\theta = 0$) thereafter. The top panel displays transient angular distributions at $t = 20$, showing lower degrees of adjustment to the new gradient direction for stronger alignment tendencies. The bottom panel displays average up-gradient velocity as a function of time. Although the equilibrium velocity is highest for the strongly aligning case, strong alignment also prevents a group from rapidly correcting course when the gradient direction changes. In this particular scenario, an intermediate tendency ($\beta = 1$) appears to result in the highest up-gradient motion.

4 Slime mold aggregation

After exhausting their food supply, the amoebae first tend to distribute themselves uniformly over the space available to them, but later they begin to aggregate in a number of centers. The point of view taken here is that early in the life cycle of the amoebae the properties of the cells are such that a uniform distribution is stable. Random non-uniformities which inevitably occur in a real system, decay. At some point in the life cycle of all the cells, however, the characteristics of the individual cell change in such a way as to make the uniform distribution unstable. Hence, the initiation of aggregation is viewed as a breakdown of stability. This point of view does not require that cells be distinguished.

Here, we derive a simplified model for the chemotactic motion of amoebae and analyze the linear stability of the uniform state. The conditions for instability illustrate the balance of effects responsible for aggregation. Linear theory will provide a spatial pattern of aggregation corresponding to the largest temporal growth rate. The pattern which is actually observed and the saturation mechanism for arresting the growth will, no doubt, depend on the nonlinear effects.

4.1 Formulation

Acrasin is the chemotactic attractant that is produced by cells. Let $a(x,y,t), \rho(x,y,t)$ and $\eta(x,y,t)$ represent the concentration (mass per unit area) of amoebae, acrasin and acrasinase respectively. Acrasin is produced by the amoebae at a rate $f(\rho)$ per amoeba (allowing for the acrasin production to depend on the ambient acrasin concentration). Acrasinase is produced by the amoebae at a rate $g(\rho, \eta)$ per amoeba (allowing for the acrasinase production to depend on the ambient acrasin and acrasinase concentrations). Acrasinase is an enzyme which breaks down acrasin by first forming a complex C (of concentration c) which dissociates into the free enzyme (acrasinase) plus the degraded product i.e.,

$$\rho + \eta \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_2}{\rightarrow} \eta + \text{product.}$$
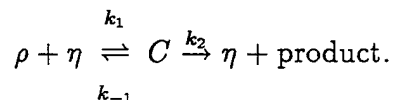
Acrasin, acrasinase and the complex undergo Fickian diffusion. The amoeba concentration changes as a result of an oriented chemotactic motion towards positive acrasin gradient and a random motion analogous to diffusion. Hence, we have

$$\partial a / \partial t = -\nabla \cdot (D_1 \nabla \rho) + \nabla \cdot (D_2 \nabla a) \tag{10}$$

$$\partial \rho / \partial t = -k_1 \rho \eta + k_{-1} c + a f(\rho) + D_\rho \nabla^2 \rho \tag{11}$$

$$\partial c / \partial t = k_1 \rho \eta - (k_{-1} + k_2) c + D_c \nabla^2 c \tag{12}$$

$$\partial \eta / \partial t = -k_1 \rho \eta + (k_{-1} + k_2) c + a g(\rho, \eta) + D_\eta \nabla^2 \eta \tag{13}$$

Here k_1, k_{-1} and k_2 are the rate constants for the acrasin-acrasinase reaction, D_ρ, D_η and D_c are constant coefficients of diffusion and D_1 and D_2 are functions of ρ and a and measure the strength of the influence of the acrasin gradient on the flow of amoebae and the vigor of the random motion of the individual amoebae respectively.

4.2   Simplification of chemistry

Notice that the amoebae respond chemically only to acrasin. The effect of acrasinase and the complex is to remove acrasin from the system. Here, we make some assumptions which allow us to model the effect of chemistry through a destruction term for acrasin which depends only on the ambient acrasin concentration. This leads to a simplified two equation model for the amoebae and acrasin concentrations.

The first is Haldane's assumption that the complex is in a steady state with regard to the chemical reaction :

$$k_1 \rho \eta - (k_{-1} + k_2)c = 0. \tag{14}$$

Next, we assume that the total concentration of enzyme (both free and bound) is a constant, η_0:

$$\eta + c = \eta_0. \tag{15}$$

Equations (14) and (15) are an alternative to (12) and (13).

It follows from (14) and (15) that

$$\eta = \eta_0/[1 + K\rho] \quad \text{where} \quad K = k_1/(k_{-1} + k_2) \tag{16}$$

With (14) and (16), (11) becomes

$$\partial \rho / \partial t = -k(\rho)\rho + af(\rho) + D_\rho \nabla^2 \rho, \tag{17}$$

where

$$k(\rho) = \eta_0 k_2 K/(1 + K\rho). \tag{18}$$

Rewriting (10)

$$\partial a / \partial t = -\nabla \cdot (D_1 \nabla \rho) + \nabla \cdot (D_2 \nabla a) \tag{19}$$

Equations (17), (18) and (19) form the simplified model for a and ρ.

4.3   Linear stability of the uniform state

A uniform state for the amoebae and acrasin prior to aggregation appears to be consistent with observation. Equations (17), (18) and (19) possess an equilibrium solution of the form

$$a = a_0, \rho = \rho_0 \tag{20}$$

where a_0 and ρ_0 are constants such that the rate of acrasin production matches the rate of acrasin destruction i.e.,

$$a_0 f(\rho_0) = k(\rho_0)\rho_0. \tag{21}$$

We perform a linearized temporal stability analysis about this state. Hence, expand

$$
\begin{align}
a(x,y,t) &= a_0 + \bar{a}(x,y,t) \tag{22}\\
\rho(x,y,t) &= \rho_0 + \bar{\rho}(x,y,t) \tag{23}
\end{align}
$$

valid for small perturbations about the uniform state. Upon substitution into (17), (18) and (19) and ignoring higher than first-order terms in the fluctuations, we obtain

$$
\begin{align}
\partial\bar{\rho}/\partial t &= -\bar{k}\bar{\rho} + a_0 f'(\rho_0)\bar{\rho} + f(\rho_0)\bar{a} + D_\rho \nabla^2 \bar{\rho} \tag{24}\\
\partial\bar{a}/\partial t &= -D_1(a_0,\rho_0)\nabla^2 \bar{\rho} + D_2(a_0,\rho_0)\nabla^2 \bar{a} \tag{25}
\end{align}
$$

where a prime denotes a derivative and

$$\bar{k} \equiv k(\rho_0) + \rho_0 k'(\rho_0). \tag{26}$$

The linearized equations (24) and (25) have coefficients which are not functions of space or time and so we seek solutions of the form

$$
\begin{align}
\bar{a} &= \hat{a}\cos(q_1 x + q_2 y)\exp(\sigma t), \tag{27}\\
\bar{\rho} &= \hat{\rho}\cos(q_1 x + q_2 y)\exp(\sigma t) \tag{28}
\end{align}
$$

Substituting (27) and (28) into (24) and (25), we obtain

$$(F - \sigma)\hat{\rho} + f(\rho_0)\hat{a} = 0, \tag{29}$$

$$D_1 q^2 \hat{\rho} - (D_2 q^2 + \sigma)\hat{a} = 0, \tag{30}$$

where it is understood that D_1 and D_2 are evaluated at $a = a_0, \rho = \rho_0$, and

$$q^2 \equiv q_1^2 + q_2^2, \quad F \equiv f'(\rho_0)a_0 - \bar{k} - q^2 D_\rho. \tag{31}$$

For a non-trivial solution, the determinant of the coefficients vanishes i.e.,

$$\sigma^2 - \sigma(F - q^2 D_2) - (q^2 f(\rho_0)D_1 + q^2 D_2 F) = 0. \tag{32}$$

This is a quadratic equation in σ of the form

$$\sigma^2 + b\sigma + c = 0. \tag{33}$$

The discriminant $(b^2 - 4c)$ is positive and hence the roots are real. The condition for stability is that both the roots be negative or $b > 0$ and $c > 0$ i.e.,

$$F < q^2 D_2, \quad F < -f(\rho_0)D_1/D_2. \tag{34}$$

107

Since $f(\rho_0), D_1$ and D_2 are positive, $F < -f(\rho_0)D_1/D_2$ is sufficient. For a given value of q, the sole condition for stability is thus

$$D_1 f(\rho_0) + D_2 f'(\rho_0)a_0 < D_2(\bar{k} + D_\rho q^2). \tag{35}$$

Instability will occur if the left side of the above equation exceeds the right side. This will first occur for disturbances associated with a zero value of q (by solving (32) one can also show that unstable disturbances with $q = 0$ always grow fastest). The instability condition is, therefore,

$$\frac{D_1 f(\rho_0)}{D_2 \bar{k}} + \frac{a_0 f'(\rho_0)}{\bar{k}} > 1. \tag{36}$$

Rewriting using (21), we have

$$\frac{D_1 \rho_0}{D_2 a_0} + \frac{a_0 f'(\rho_0)}{\bar{k}} > 1. \tag{37}$$

Here, we discuss the physical mechanisms responsible for aggregation as indicated by the instability condition (36).

Case 1 : $D_1 f(\rho_0) > D_2 \bar{k}$

This is the only effect in the event that acrasin production is independent of the ambient concentration of acrasin (i.e., $f(\rho)$ a constant). Suppose there is a high concentration of amoebae and acrasin initially at some point P. The greater the equilibrium acrasin production rate, $f(\rho_0)$, the more rapidly the local concentration of acrasin-producing amoebae at P raises the acrasin level still further. In addition, a large value of D_1 will cause the amoebae near P to be more strongly attracted by the relatively high concentration of acrasin at P. Hence, a slight perturbation can reinforce itself. On the other hand, a large value of D_2 indicates a strong tendency to smooth out the local maximum in amoeba density by "diffusion". Similarly, a large value of the decay coefficient $\bar{k}$, will more effectively flatten out the local maximum in acrasin density. When $a_0 f'(\rho_0)/\bar{k}$ is negligible compared to one, it is thus the relative predominance of the combined effects of diffusion and decay on the one hand, and acrasin production and the chemotactic response on the other hand, which determine the stability or instability of a uniform distribution of amoebae and acrasin.

Case 2 : $a_0 f'(\rho_0) > \bar{k}$

This implies that a small increase in acrasin level causes an increase in acrasin output which outweighs its more rapid decay. This would lead immediately to an instability in the level of acrasin. This effect cannot operate in isolation of the previous one as that would require that $D_1 = 0$ which would make the amoeba not responsive to acrasin, and aggregation would not be observed.

4.4 Discussion

What does perturbation with $q = 0$ mean ?

If territory size is identified with the wavelength of the most unstable disturbance, then the theory predicts infinite territory sizes. Really what the prediction $q = 0$ is saying is that the wavelength at the onset of aggregation is so large that it cannot be quantified in this somewhat crude model.

Effect of threshold in the chemotactic response

Let us assume that the amoebae are not sensitive to acrasin gradients unless the spatial average of $|\nabla \rho / \rho|$ is larger than some constant α. The instability condition is modified to

$$\frac{D_1 f(\rho_0)}{D_2 \bar{k}} + \frac{a_0 f'(\rho_0)}{\bar{k}} > 1 + D_\rho \frac{\alpha^2 \rho_0^2}{\hat{\rho}^2}. \tag{38}$$

Thresholding is believed to have the effect of limiting territory sizes. Because an amoeba sensing the local gradient can move up a gradient even if the signal to noise ratio was fairly low, it is unclear whether it makes sense to threshold the chemotactic response.

What properties are not necessary for aggregation ?

Encouraging agreement between the experiments and the predictions made by this theory shows that the following properties are unnecessary to explain the onset of aggregation and the non-random spacing of centers : (a) markedly special properties of the centers; (b) acquisition of stickiness; (c) abrupt changes in cellular physiology; (d) orienting pulses paced by the center; (e) a spacing substance.

References

[1] ALT, W. 1980 Biased random walk models for chemotaxis and related diffusion approximations. *J. Math. Biol.* **9**, 147-177.

[2] DAVIS, C. S., FLIERL, G. R., WIEBE, P. H. & FRANKS, P. J. S. 1991 Micropatchiness, turbulence and recruitment in plankton. *J. Mar. Res.* **49**, 1-43.

[3] GRÜNBAUM, D. & OKUBO, A. 1994 Modeling social animal aggregations. *Frontiers in Theoretical Biology (ed. S. A. Levin)* Lecture Notes in Biomathematics, vol. 100, Springer-Verlag.

[4] GRÜNBAUM, D 199x *Schooling as a strategy for chemotaxis in a noisy environment.* to appear.

[5] KELLER, E. F. & SEGEL, L. A. 1971 Model for chemotaxis. *J. Theor. Biol.* **30**, 225-234.

[6] LEVIN, S. A. 1986 Random walk models and their implications. *Mathematical Ecology (eds. T. G. Hallam & S. A. Levin)* Biomathematics, vol. 17, Springer-Verlag.

[7] MURRAY, J. D. 1993 *Mathematical Biology* Biomathematics, vol. 1, Springer-Verlag.

[8] OKUBO, A. 1980 *Diffusion and Ecological Problems : Mathematical Models* Biomathematics, vol. 10, Springer-Verlag.

[9] OKUBO, A. 1986 Dynamical aspects of animal grouping : swarms, schools, flocks, and herds. *Advances in Biophysics* **22**, 1-94.

[10] OTHMER, H. G., DUNBAR, S. R. & ALT, W. 1988 Models of dispersal in biological systems. *J. Math. Biol.* **26**, 263-298.

Notes compiled by Scott Stewart and Amar Gandhi.

Lecture IX: Glenn Flierl – The Physics of Patchiness
Example 1: A patch in a shear flow

Let the density of a certain biological population be described by the Fokker-Planck equation:

$$\frac{Dn}{Dt} = \nabla^2 \left(K(x,n) \cdot n \right) \tag{1}$$

and let K be described by an areal integral of a weighting function multiplied by the concentration of organisms:

$$K(\bar{x}, n) = F \left(\int\int w(x, x')n(x')dx \right)$$

where w is a monotonically decreasing function of the distance $(x - x')$, and F decreases monotonically with the weighted density, n. We require that

$$\int\int w(x, x')dx = 1 \quad .$$

An example: Let us impose a simple shear flow field: $\bar{u} = (u, v) = (Sy, 0)$, and choose a simple weighting function:

$$w(x, x') = \frac{\gamma^2}{2\pi} K_0 \left(\gamma r \right) \quad . \tag{2}$$

Defining m by:

$$m \equiv \int\int w(x, x')n(x')dx \quad ,$$

and using the fact that w is $-\gamma^2$ times the Green's function G which satisfies

$$\left(\nabla^2 - \gamma^2 \right) G = \delta(r) \quad , \tag{3}$$

we get

$$\left(\nabla^2 - \gamma^2 \right) m = -\gamma^2 n \tag{4}$$

111

which is coupled to equation (1):

$$\frac{Dn}{Dt} = \nabla^2 \left(F(m) \cdot n \right) \quad . \tag{5}$$

Denoting spatial averages by overbars, we notice from (4) that $\bar{m} = \bar{n}$, since $\overline{\nabla^2 m} = 0$. Spatially averaging equation (5), we also see that $\bar{n}$ is an equilibrium solution, since $\frac{D\bar{n}}{Dt} = 0$. Letting primes denote the deviation from the spatial average, we can linearize equation (5) around the equilibrium solution

$$\left(\frac{\partial}{\partial t} + Sy\frac{\partial}{\partial k} \right) n' = \nabla^2 \left(F(\bar{m}) \cdot n' + \frac{dF(\bar{m})}{dm}\bar{n}m' \right) \quad .$$

Assuming a normal mode solution in space and a growing mode in time, $n \propto e^{\sigma t} e^{ikx}$, we find

$$(\sigma + Syik)n' = \nabla^2 \left(F(\bar{m}) \cdot n' + \frac{dF(\bar{m})}{dm}\bar{n}m' \right) \tag{6}$$

and

$$\left(k^2 - \frac{\partial^2}{\partial y^2} + \gamma^2 \right) m' = \gamma^2 n' \quad .$$

Solving for the special case $k \equiv 0$ and $\frac{\partial}{\partial y} \rightarrow il$, we find that

$$m' = \frac{\gamma^2}{l^2 + \gamma^2}n'$$

which can be plugged into equation (6) to give

$$\sigma = -l^2 \left(F(\bar{m}) + \frac{dF(\bar{m})}{dm}\bar{n}\frac{\gamma^2}{l^2 + \gamma^2} \right) \quad .$$

From this solution, we can see that an instability will develop if the term in brackets is negative. For the particular choice of F

$$F(\bar{m}) = e^{-\lambda \bar{m}} \quad ,$$

an instability will develop if

$$\lambda > \left(\frac{\gamma^2 + l^2}{\gamma^2} \right) \frac{1}{\bar{n}} \quad .$$

112

With growth in the band $0 < l^2 < \gamma^2(\lambda \bar{n} - 1)$, this solution indicates that a sufficiently dense population can group into bands oriented along the shear. In the absence of shear, the bands can be arranged in any direction and we might expect the patch distribution to be isotropic. Equation (6) suggests that shear may break up patches which vary in x.

Numerical experiments with equations (4) and (5) shown in Figure 1 demonstrate the evolution of patches and the alignment in the presence of shear.

Example 2: Circulation and biology around Georges Bank

We can model the (steady) flow past Georges Banks as flow in a rectangular region, with solid northern and western boundaries, and constant inflow along the eastern boundary. The bank is represented as a circular bump (Figure 2).

The equation for conservation of potential vorticity in this inviscid system is:

$$\frac{D}{Dt}\left(\frac{\zeta + f}{h}\right) = 0 \quad . \tag{7}$$

where $h(\mathbf{x})$ is the depth of the fluid at $\mathbf{x}$, ζ is the vertical component of the relative vorticity, defined by

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \quad ,$$

and f is the Coriolis parameter, approximated by

$$f = f_0 + \beta y \quad .$$

For the steady state or rigid lid system, the conservation of mass equation is simply

$$\nabla \cdot (\bar{u}h) = 0 \quad ,$$
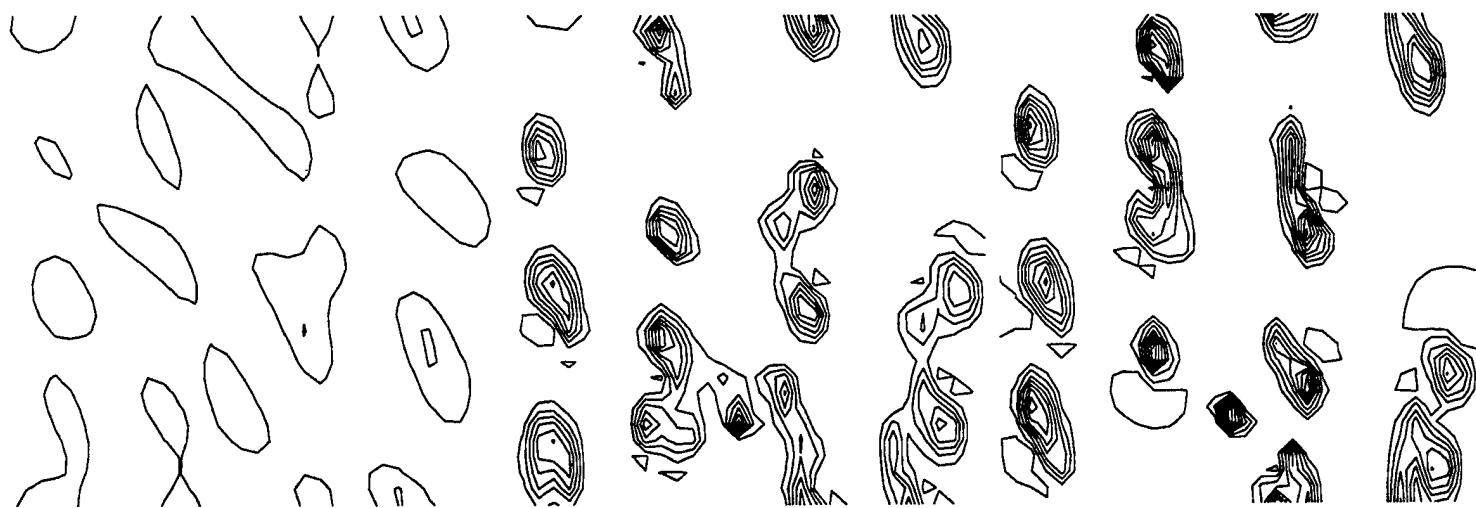
so we can write the velocity in terms of a stream function, ψ:

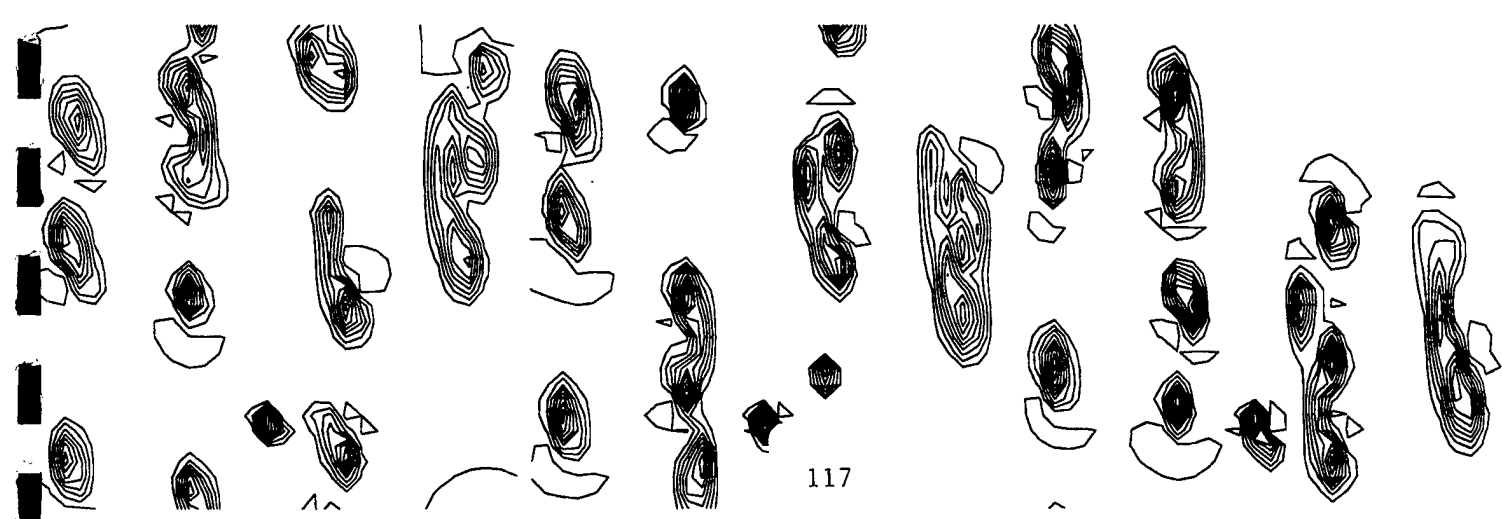$$uh = -H\psi_y \quad \text{and} \quad vh = H\psi_x \quad .$$

S = 0

114

S=0.1

116
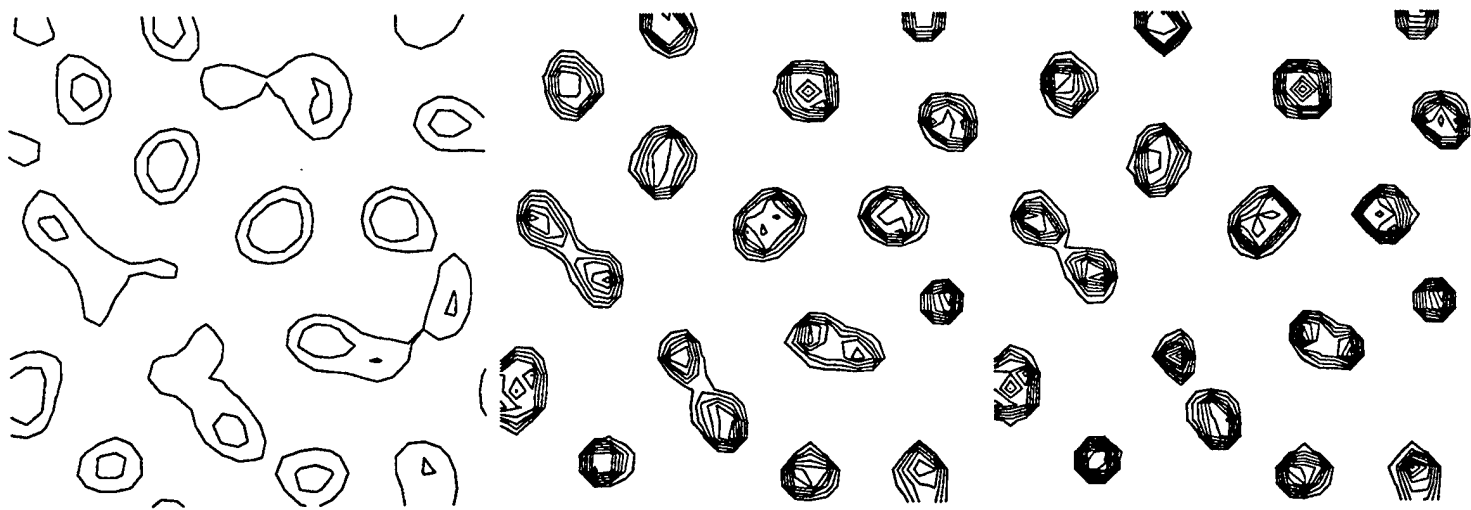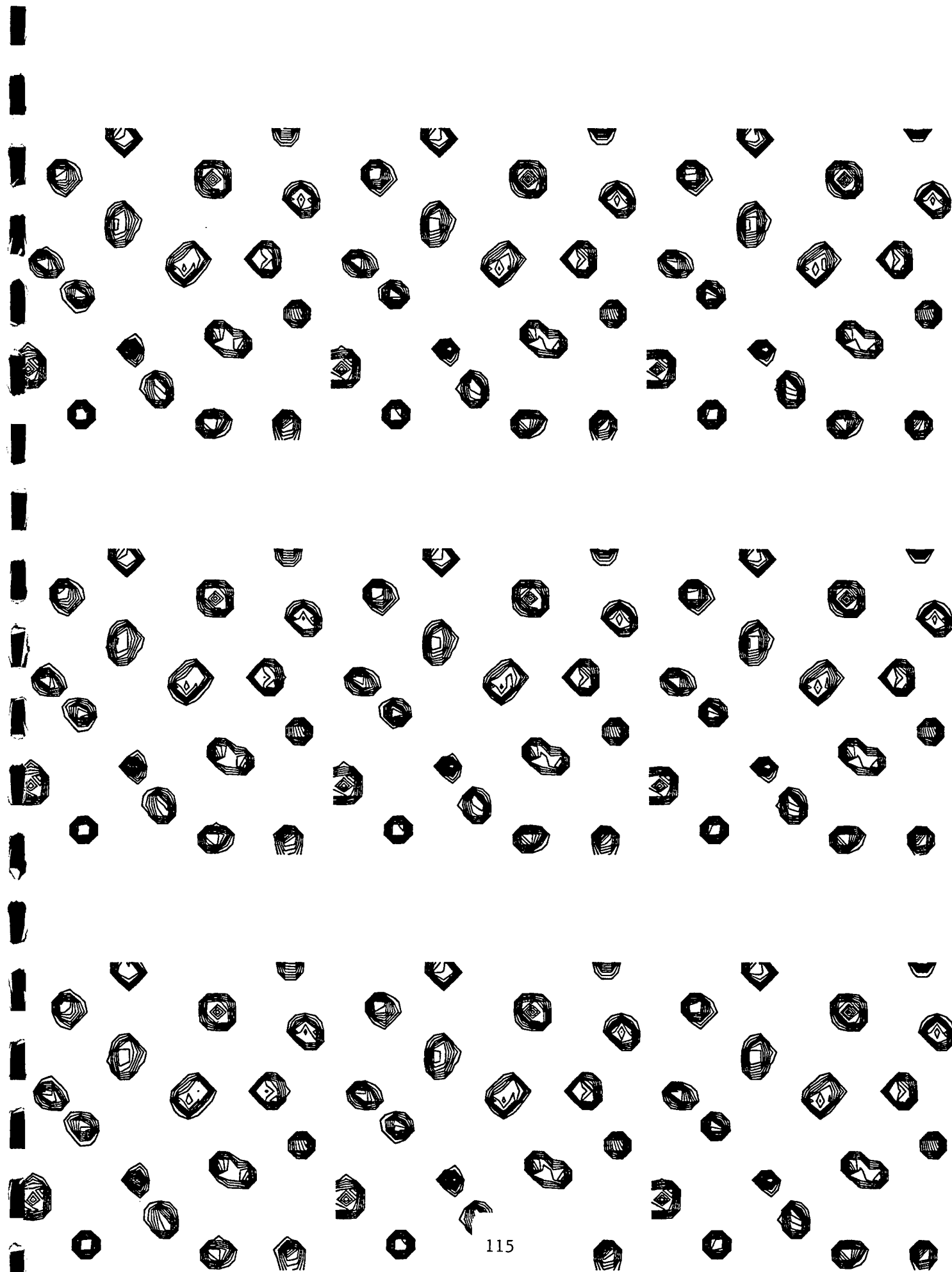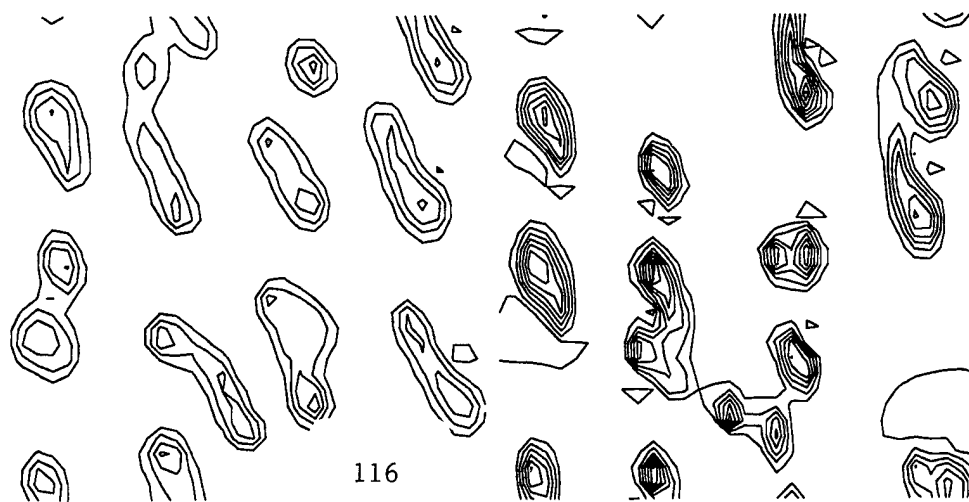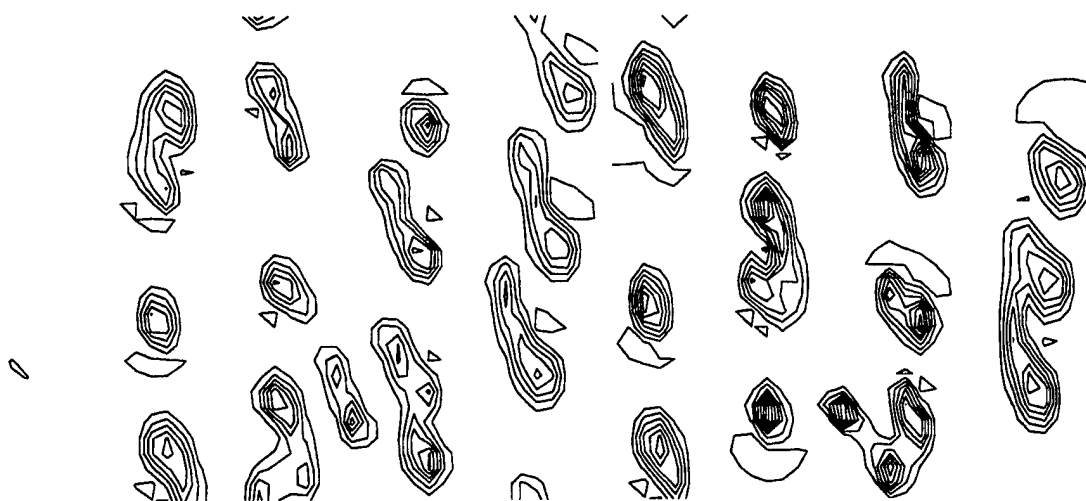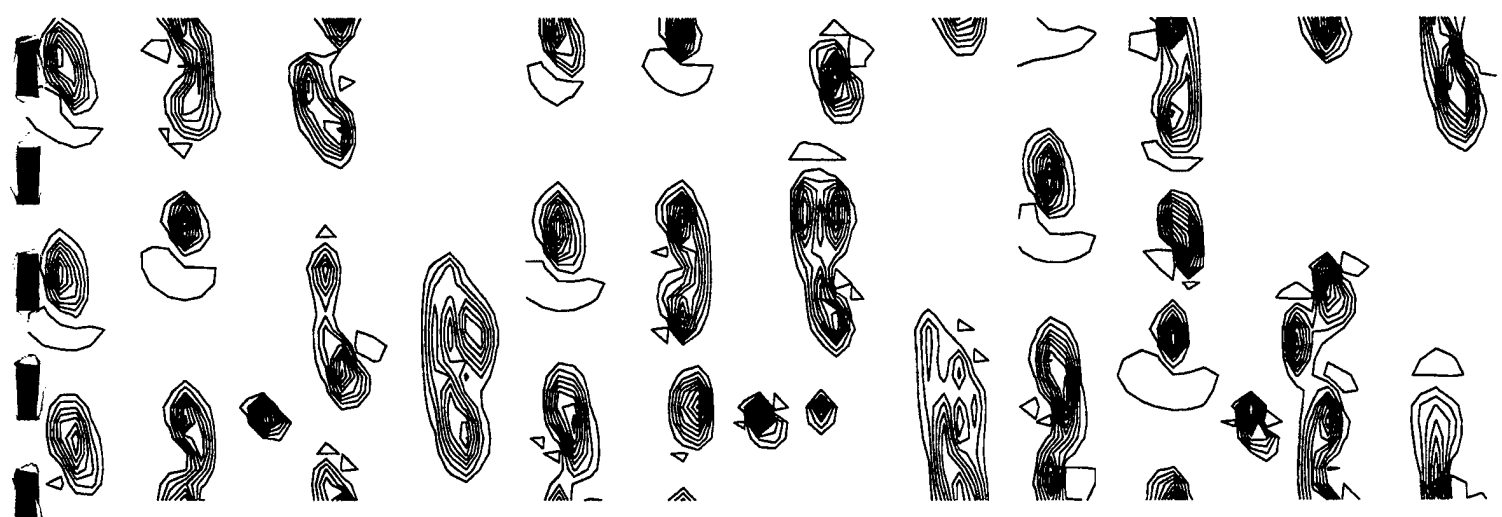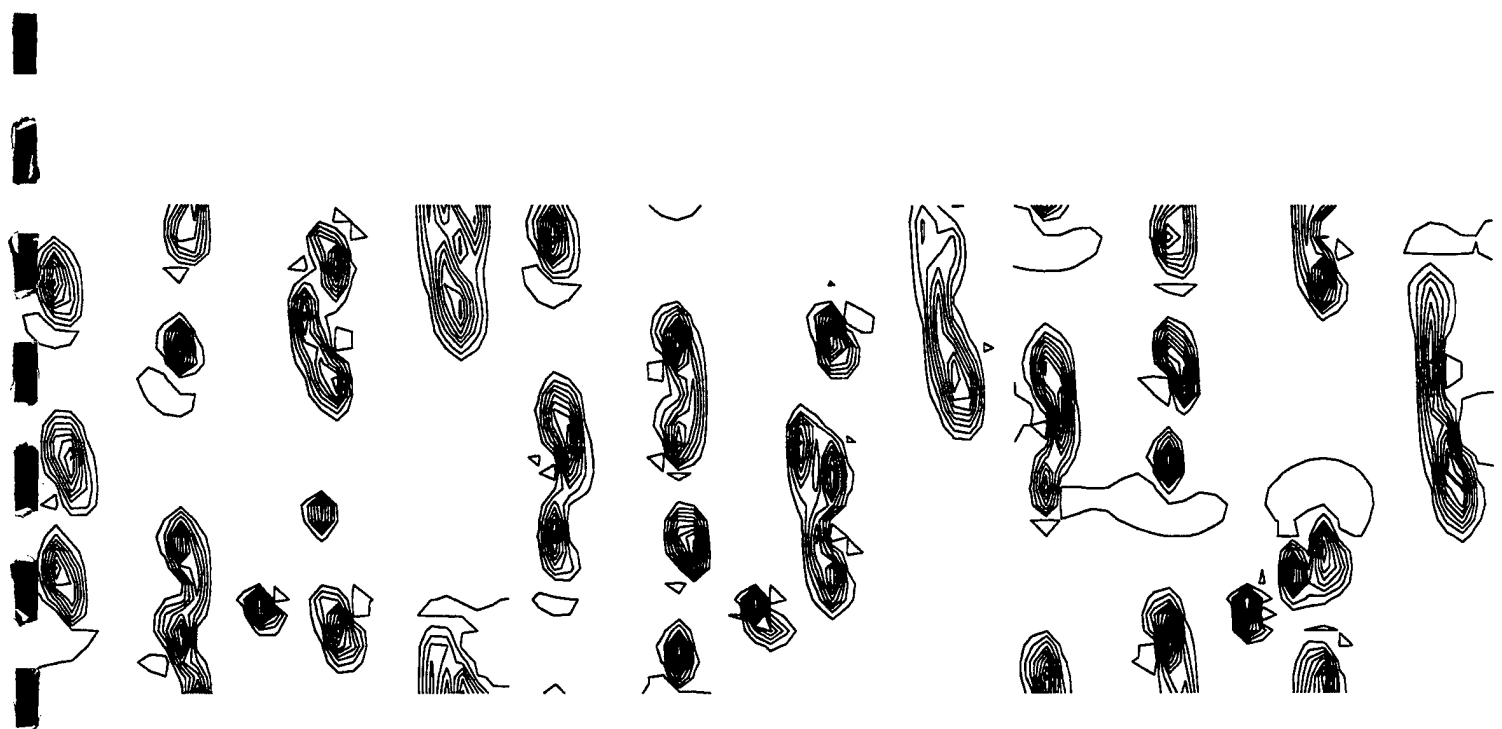
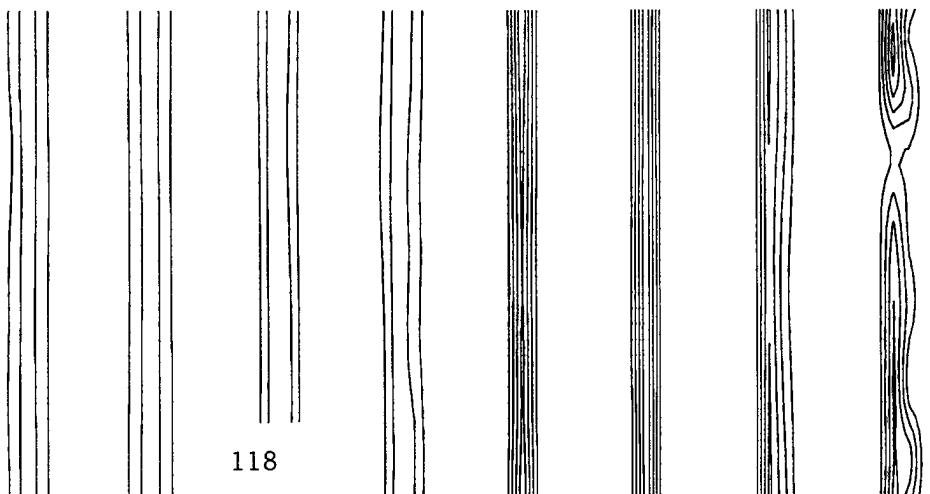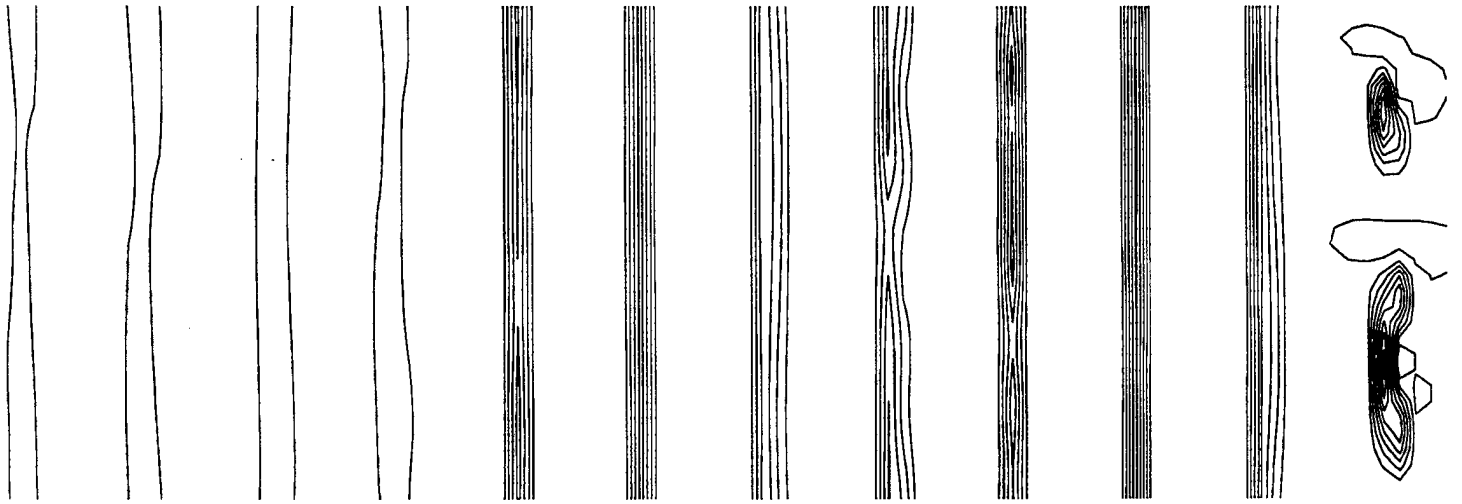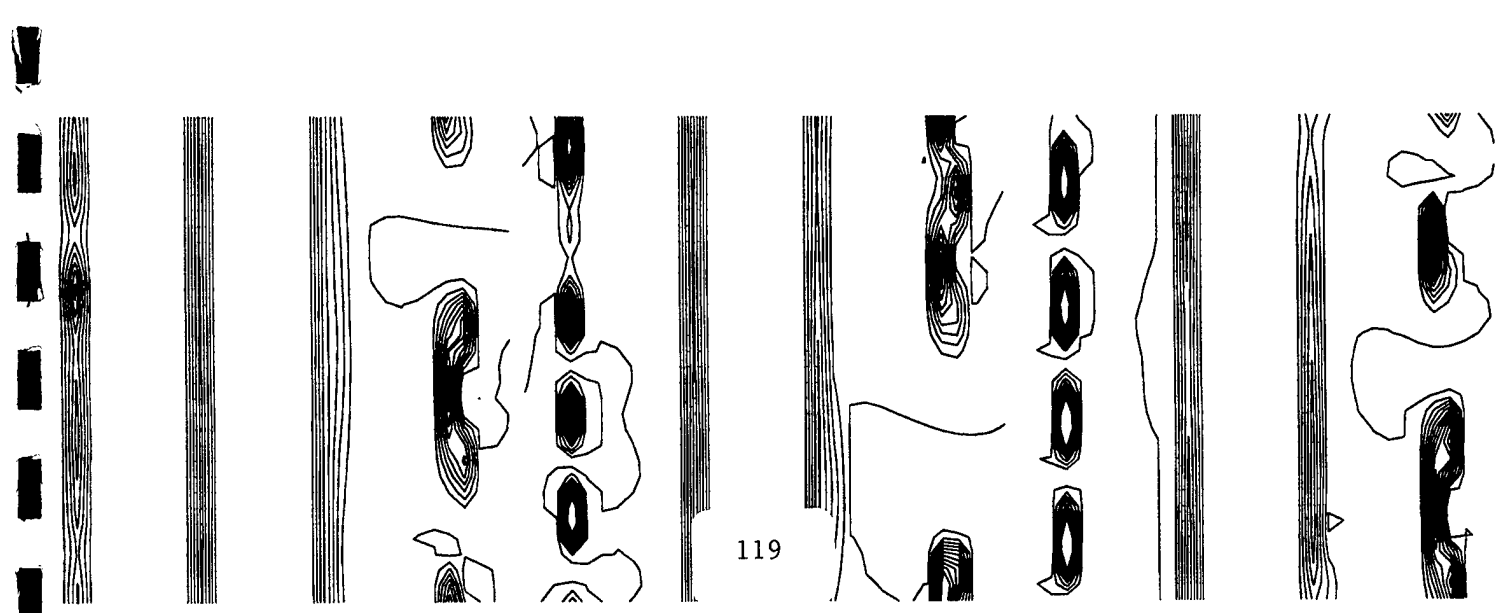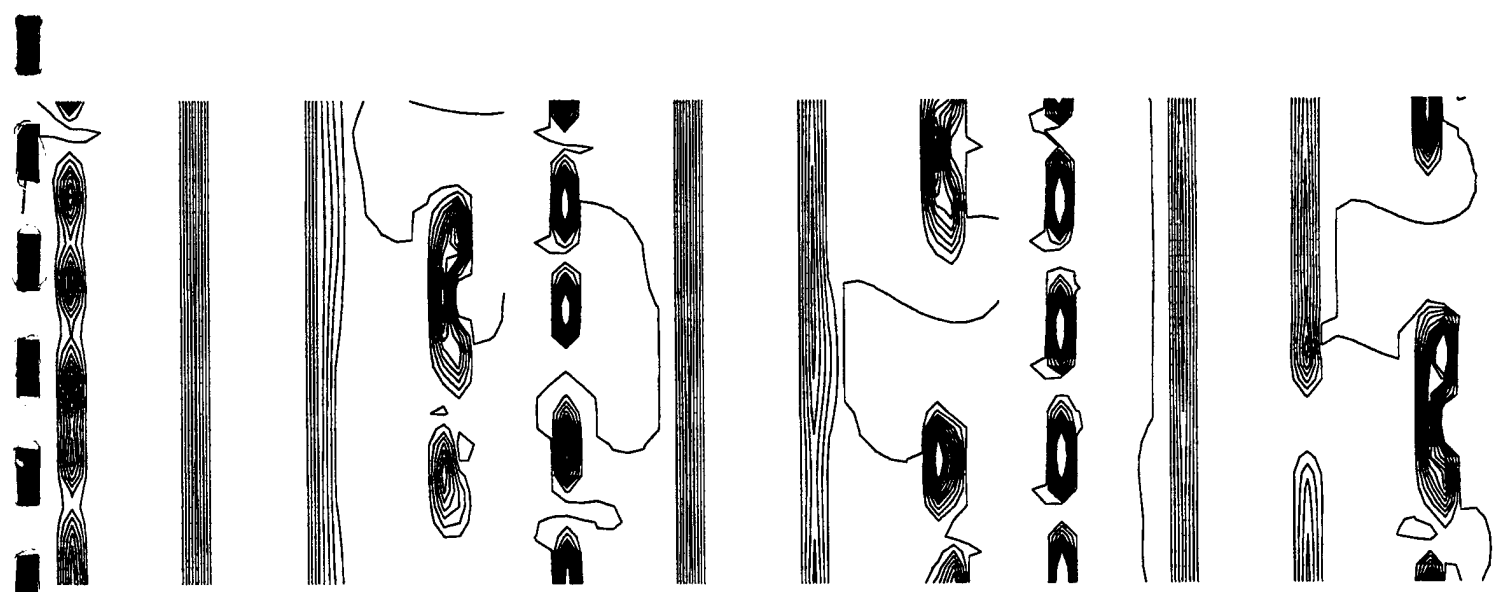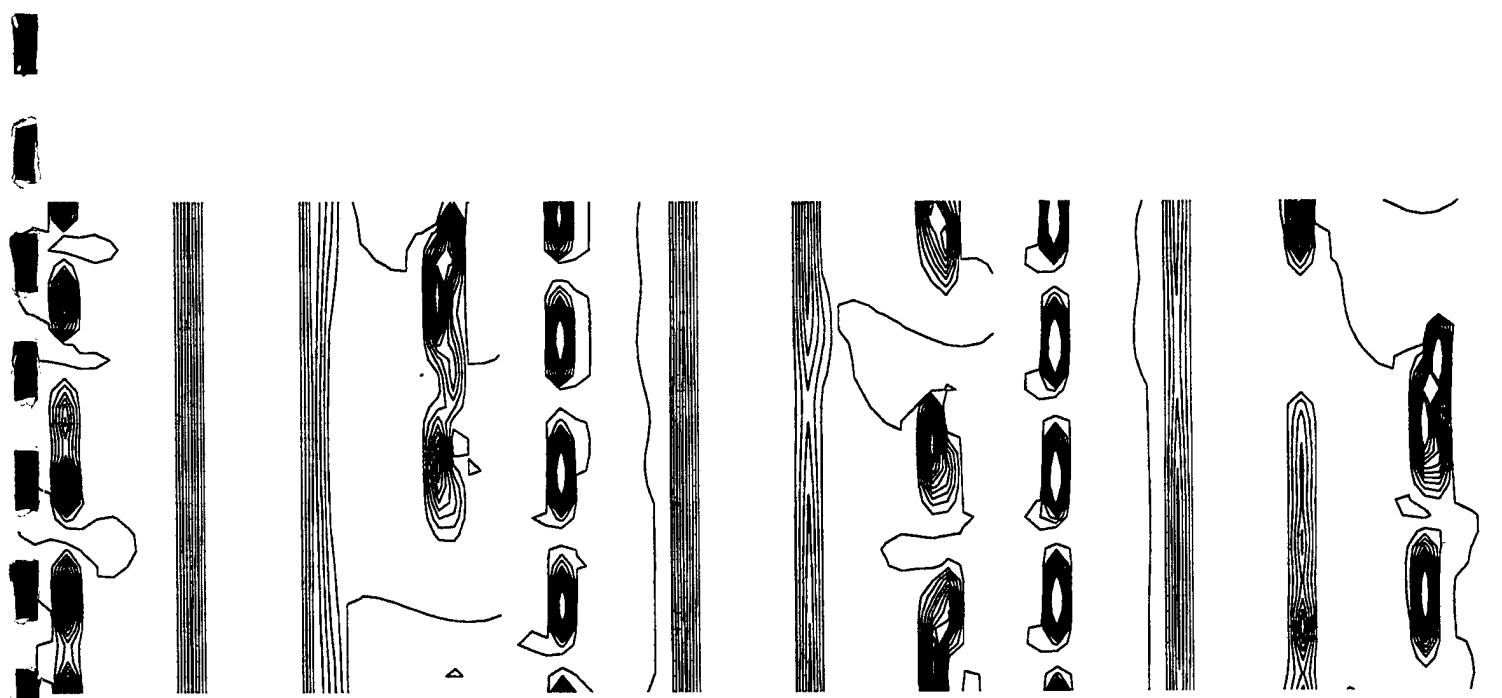$S = 0.5$
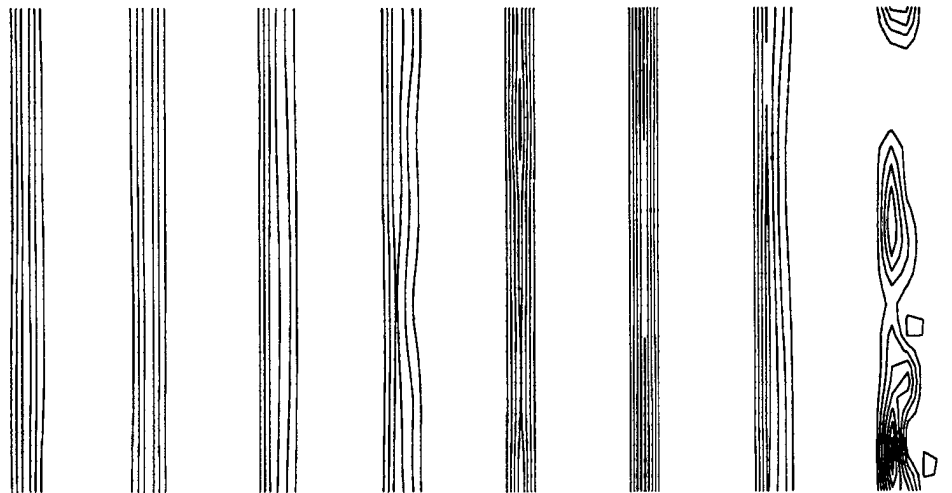
Letting q be the potential vorticity

$$q = \frac{\nabla \cdot \frac{H}{h} \nabla \psi + f_0 + \beta y}{h} \quad , \tag{8}$$

we can rewrite equation (7) as follows

$$\left(\psi_x \frac{\partial}{\partial y} - \psi_y \frac{\partial}{\partial x} \right) q = 0$$

which indicates that q is constant along streamlines.

At the eastern boundary, both the depth and velocity are constant, with $h = H$, and $u = U_0$. The stream function and potential vorticity in this region are therefore

$$\psi = -U_0 y \quad \text{and} \quad q = \frac{f_0 + \beta y}{H} = \frac{1}{H} \left(f_0 - \frac{\beta}{U_0} \psi \right) \quad .$$

Because the potential vorticity is conserved along streamlines, this solution for q must hold at all points in the domain. (This is not strictly true for regions where the streamlines are closed, but we will assume that q has the same form along closed streamlines.) Plugging this solution for q into equation (8) gives an equation for ψ everywhere in the domain:

$$\nabla \cdot \frac{H}{h} \nabla \psi + f_0 + \beta y = \frac{h}{H} \left(f_0 - \frac{\beta}{U_0} \psi \right) \quad ,$$

or, if we assume that the height of the bottom topography is small compared with the maximum depth (i.e. that $\frac{h}{H} \approx 1$),

$$\left(\nabla^2 + \frac{\beta}{U_0} \right) \psi = -\frac{H - h}{H} f_0 - \beta y \quad . \tag{9}$$

Solving first for the inertial boundary current away from the bank (where $h = H$), we find that

$$\psi_1 = -U_0 y \left(1 - \exp \left(-\sqrt{\frac{-\beta}{U_0}} x \right) \right)$$

or, if $\frac{\beta}{U_0} \equiv -k^2$

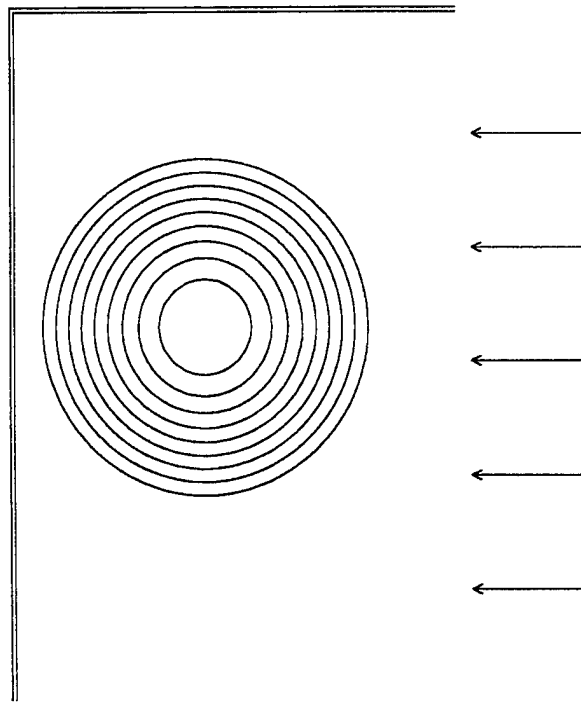$$\psi_1 = -U_0 y \left(1 - e^{-kx} \right) \quad .$$

Figure 2: Topography, showing shallow bank in rectangular region with uniform inflow
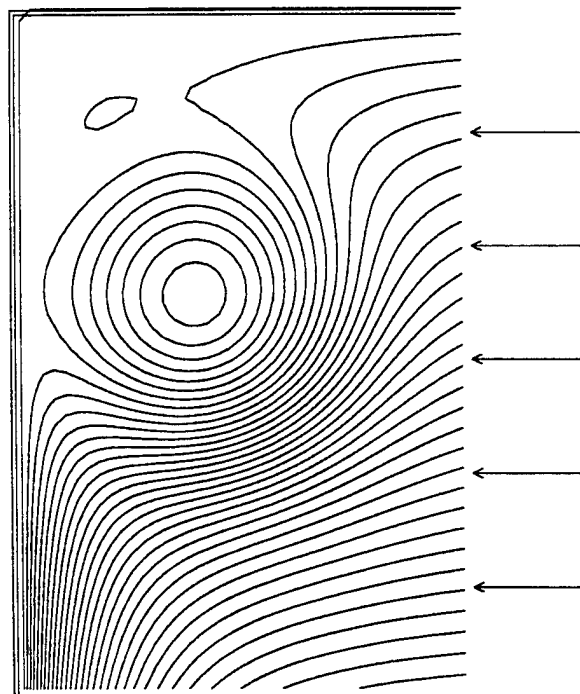

Figure 3: Streamfunction for the flow

In the region of the bank, (where h decreases,) ζ must also decrease to conserve potential vorticity, resulting in clockwise (anticyclonic) circulation around the bank. The equation for ψ_2 in this region is

$$\left(\nabla^2 - k^2\right)\psi_2 = -\frac{H-h}{H}f_0$$

with the boundary condition that ψ_2 goes to zero on the eastern boundary. This equation can be solved with Green's functions and the method of images.

The combined solution for the stream function $(\psi_1 + \psi_2)$ is shown in figure 3. (Note: for this figure, the scale of the bank was taken to be twice the characteristic width of the inertial boundary current – i.e. $r_0 = 2k$)

Now, we can add some simple biology to this model....

The biological model follows an unstructured population of organisms (B) described by the following equation:

$$\frac{DB}{Dt} = \left[\frac{\partial}{\partial t} + \left(\psi_x \frac{\partial}{\partial y} - \psi_y \frac{\partial}{\partial x}\right)\right]B = g(h)B - \alpha B^2 + K\nabla^2 B \tag{10}$$

where $g(h)$ is the growth function, αB^2 represents logistic interactions, and $K\nabla^2 B$ represents the horizontal diffusion of the organisms. The growth of the organisms is assumed to be a function of depth, because the region is relatively shallow, and the mixed layer extends over the full depth.

Solutions:

I. In one dimension, if the velocity is zero and the growth is assumed constant, equation (10) simplifies to Fisher's equation

$$\frac{\partial B}{\partial t} = g(h)B - \alpha B^2 + K\frac{\partial^2 B}{\partial x^2} \tag{11}$$

which has two equilibrium solutions

$$B = 0 \quad \text{and} \quad B = \frac{g}{\alpha} \quad .$$

The unsteady solution connecting these two equilibria takes the form of a moving diffusion-growth front. We assume the following form for the solution:

$$B = B(x - ct) \quad .$$

Plugging this solution into equation (11) gives

$$-c\frac{dB}{dx} = gB - \alpha B^2 + K\frac{d^2B}{dx^2}$$

which has a solution

$$B = \frac{g}{\alpha}\left[\frac{1}{4} + \frac{1}{4}\tanh^2\frac{x}{w} - \frac{1}{2}\tanh\frac{x}{w}\right]$$

in which w is the length scale of the front:

$$w = 2\sqrt{\frac{6K}{g}} \quad .$$

The speed at which the front propagates is then given by

$$c = \frac{5}{\sqrt{6}}\sqrt{gK} \quad .$$

This is quite close to the minimum speed ($c = 2$) which tends to be the solution developing from a step function (or very sharp) initial condition (c.f., Murray, 1990).

II. There is also a straightforward solution for the one dimensional case with non-zero velocity and growth. First, using the following transformation:

$$B = S\exp\left(\int \frac{u}{2K}\right)$$

we can rewrite (10) as

$$\frac{\partial S}{\partial t} = \left(g(h) + \frac{ux}{2} - \frac{u^2}{4k}\right)S + K\frac{\partial^2 S}{\partial x^2} \quad .$$

Assuming the solution has the form

$$S = \phi(x)e^{\sigma t} \quad ,$$

we find that

$$K\frac{\partial^2 \phi}{\partial x^2} = -\left(g + \frac{ux}{2} - \frac{u^2}{4k} - \sigma\right)\phi$$

— Schroedinger's equation. The growing modes correspond to bound states.

We can model the growth of the biology over Georges Bank using

$$g = g_0\left(1 - \frac{x^2}{r_0^2}\right) \tag{12}$$

where r_0 is the half-width of the (one-dimensional) bank. In this form, the growth equation gives maximum growth over the bank, where the mixed layer depth is shallow, and light is more available.

Taking u constant, we can integrate equation (12) to give

$$\phi = \exp\left(\frac{-x^2}{l^2}\right) \quad,$$

where $l^2 = 2\sqrt{\frac{K}{g_0}}r_0$.

Also, solving for σ gives

$$\sigma = g_0 - \frac{u^2}{2K} - \frac{\sqrt{Kg}}{r_0}$$

which indicates that the population on the bank can only grow if u is small relative to g_0 and K. Specifically, population growth is not possible if

$$u^2 > 2K\left(g_0 - \frac{\sqrt{Kg_0}}{r_0}\right) \quad, \quad \text{or if} \quad r_0 < \sqrt{\frac{K}{g_0}} \quad.$$

The characteristic speed, c, and scale, w, in Fisher's equation for the same parameter set are

$$c \approx \sqrt{Kg_0} \quad, \quad \text{and} \quad w = \sqrt{\frac{K}{g_0}}$$

so that the flow speed must be smaller than the propagation speed of a diffusive–growth front and the bank scale must be larger than the characteristic front width in order for growth to occur.

The final solution for B is then

$$B = \exp\left(\int \frac{u}{2K}\right) \phi e^{\sigma t} \quad .$$

The first exponential in this equation acts to shift the peak in the population of organisms downstream from the region of peak growth.

III. In two dimensions, if we assume that u is large relative to K and g, the biology will predominantly be confined to the region over the bank, where the streamlines are closed. In this case, we can find a first order solution, B_0, that satisfies a simplified version of equation (10)

$$\psi_x B_y - \psi_y B_x = J(\psi, B) = 0 \quad ,$$

which indicates that B_0 is purely a function of ψ. The corresponding second order equation is then

$$\frac{\partial B_0}{\partial t} + J(\psi, B_1) = gB_0 - \alpha B_0^2 + K\nabla^2 B_0 \quad .$$

We can eliminate B_1 from this equation by integrating over the region between two closed streamlines. The resulting equation is

$$A\frac{\partial B_0}{\partial t} = A\left(\frac{\int\int g}{A}\right) B_0 - A\alpha B_0^2 + \oint_{outer} K\nabla B_0 \cdot \hat{\mathbf{n}} - \oint_{inner} K\nabla B_0 \cdot \hat{\mathbf{n}} \quad .$$

Dividing by A and defining average growth and diffusivity parameters, we can rewrite this equation as

$$\frac{\partial B_0}{\partial t} = \bar{g}B_0 - \alpha B_0^2 + \frac{\partial}{\partial \psi}\bar{K}\frac{\partial B}{\partial \psi}$$

where

$$\bar{f} = \frac{\oint \frac{f}{|\nabla \psi|} dS}{\oint \frac{1}{|\nabla \psi|} dS} \quad .$$

The numerical solution for this two-dimensional case is shown in figures 4, 5, and 6. A similar case with a source of biology in the northwest corner of the domain is shown in figure 7. The sepatrix and stagnation points may play a significant role, since $\bar{g}$ limits to the value of g at the stagnation point, while the diffusivity $\bar{K}$ vanishes to this order (since the circulation time is infinite). Matching to the exterior can be rather tricky.

125
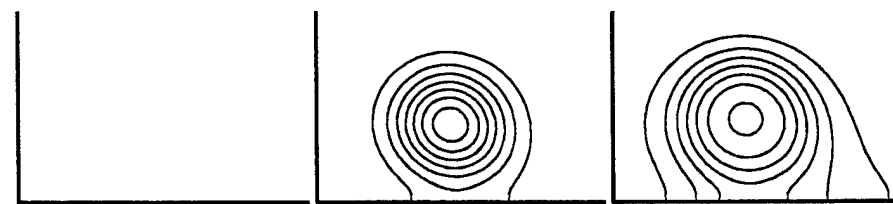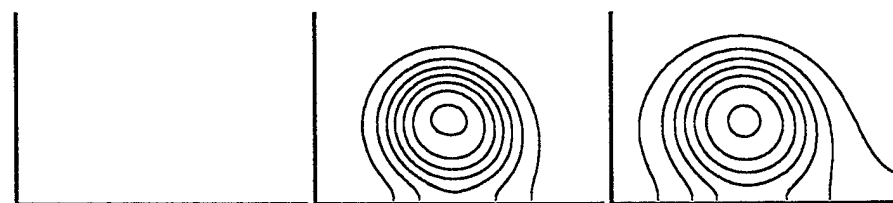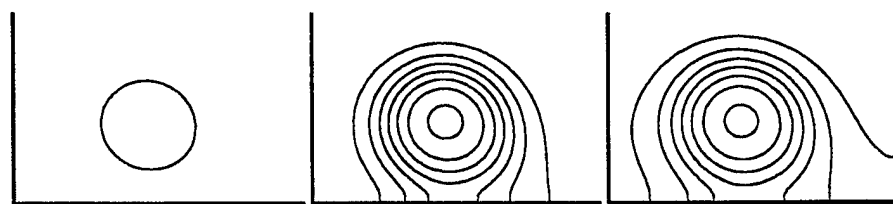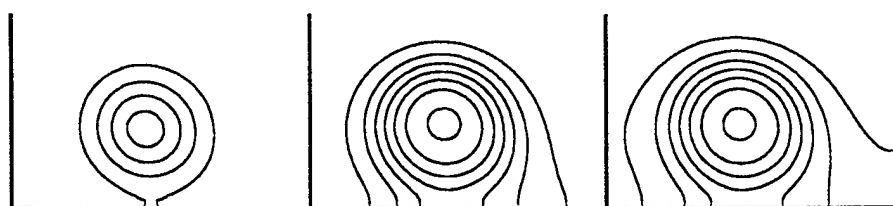
r = 0.1 to 1.3, k = 0.1

$r = -1.0$ to 0.3, $k = 0.01$

127
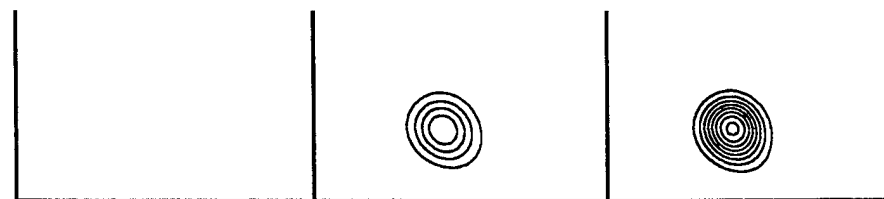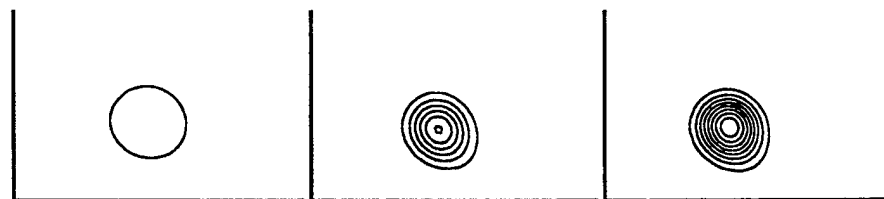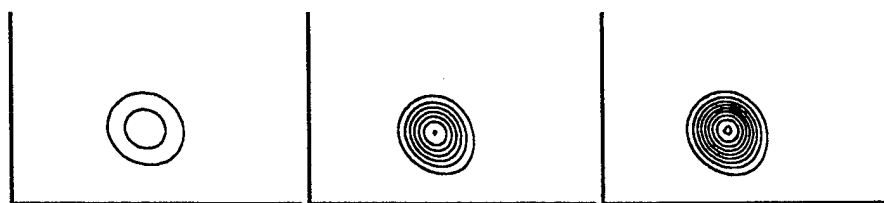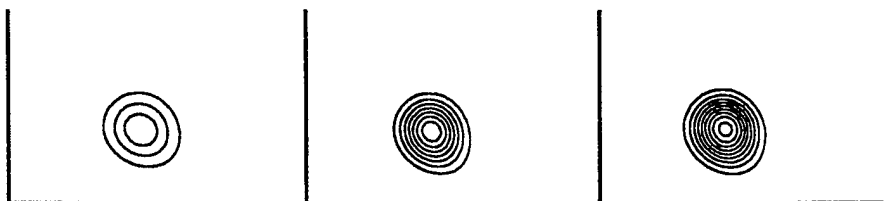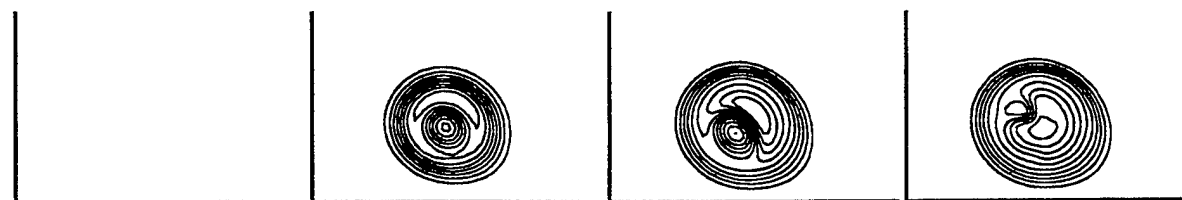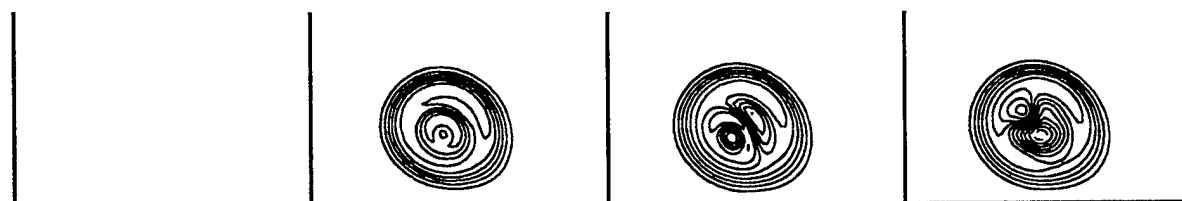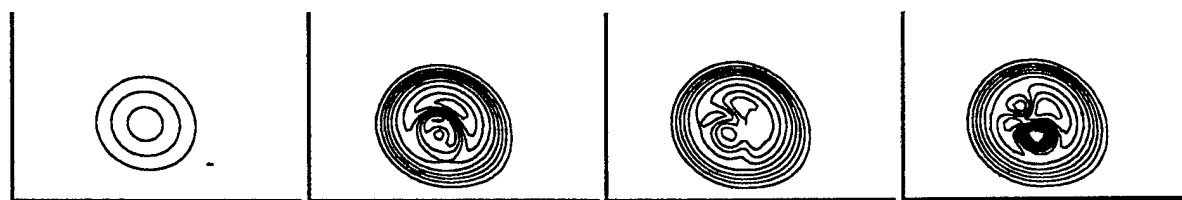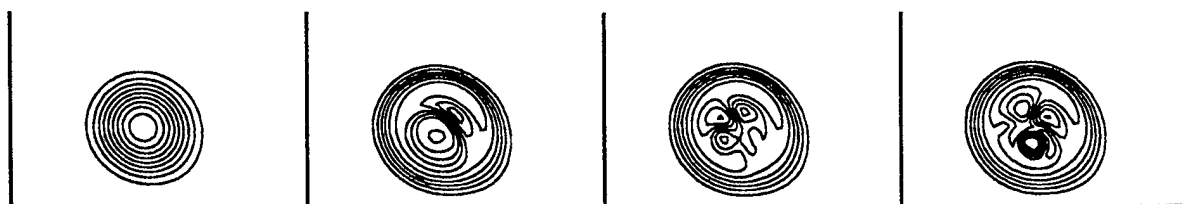
$r = -1.0$ to 1.5, $k = 0.01$
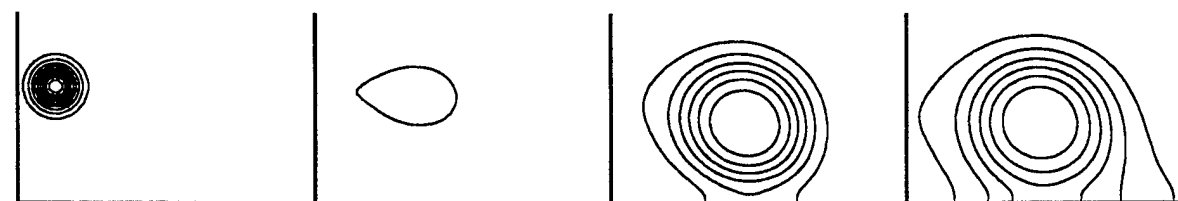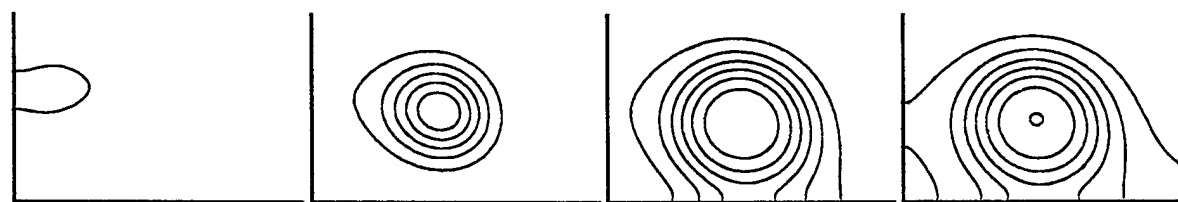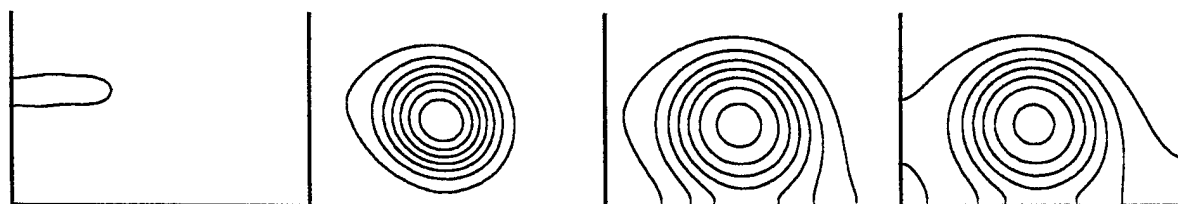
r = 0.1 to 1.3, k = 0.1

The problem of pattern and scale in ecology
Simon Levin, July 8 & 11, 1994

Credits :

Professor Levin layed out a mosaic of interesting problems in ecology and evolutionary biology which are amenable to mathematics. These lecture notes rely heavily on Prof. Levin's papers — Levin (1992), Gueron & Levin (1995), Iwasa & Levin (1995), Gueron & Levin (1993) and Durrett & Levin (1994); entire paragraphs have been lifted word for word from these papers in the writing of §1 − 5. The material reported in §6 is original work of the author of these notes.

1   Introduction

It is argued that the problem of pattern and scale is the central problem in ecology, unifying population biology and ecosystems science, and marrying basic and applied ecology (Levin, 1992). Applied challenges, such as the prediction of the ecological causes and consequences of global climate change, require the interfacing of phenomena that occur on very different scales of space, time, and ecological organization. For example, the general circulation models that provide the basis for climate prediction operate on spatial and temporal scales many orders of magnitude greater than the scales at which ecological studies are carried out. It is crucial to learn to interface the phenomena occurring at different levels.

Furthermore, there is no single natural scale at which ecological phenomena should be studied; systems generally show characteristic variability on a range of spatial, temporal, and organizational scales. The observer imposes a perceptual bias, a filter through which the system is viewed. This has fundamental evolutionary significance, since every organism is an "observer" of the environment, and life history adaptations such as dispersal and dormancy alter the perceptual scales of the species, and the observed variability. It likewise has fundamental significance for our own study of ecological systems, since the patterns that are unique to any range of scales will have unique causes and biological consequences.

The key to prediction and understanding lies in the elucidation of mechanisms underlying observed patterns. Typically, these mechanisms operate at different scales than those on which the patterns are observed; in some cases, the patterns must be understood as emerging from the collective behaviors of large ensembles of smaller scale units. In other cases, the pattern is imposed by large scale constraints. Examination of such phenomena requires the study of how pattern and variability change with the scale of description, and development of laws for simplification, aggregation, and scaling.

In this lecture, an attempt is made through four examples to address the theme of relating phenomena occurring at different levels of description. The first example looks at group size distributions emerging from social dynamics (occurring at individual level). The second

130

example focuses on temporal patterns, specifically, the timing of life history events in response to environmental conditions as selected by evolution. The third example tries to explain the striking spatial patterns of wildebeest herd fronts as an emergent property of the rules followed by individual animals whose view of the world is far smaller than the observed large-scale patterns. The fourth example uses extensions of the hawk-dove game to study effects of discreteness and spatial structure at the micro-level on macro-level behavior. Finally, we discuss connections with similar efforts made in physics (condensed matter, fluid dynamics) to relate properties at different levels of description.

2 The dynamics of group formation — Fission/Fusion

In this section, an attempt is made to relate group size distributions to processes of aggregation (fusion) and splitting (fission). Both of these processes can be related to individual behavior. For example, the decision of an individual to join or not join a group of size n, as well as the decision by individuals in a group of size n to permit or not permit an additional individual to join, may be represented in the fusion rate of groups of size 1 and groups of size n. A central question is to see whether modal group size is an epiphenomenon, not one preferred or even perceived by individuals.

Below, we derive an integro-differential balance equation for the time evolution of a group size distribution. Setting the time derivative in this equation to zero yields the equation that the stationary distribution must satisfy. Explicit solutions are found when the fission/fusion scenario is simplified. Resulting stationary distributions are discussed.

Generality of approach is sacrificed here to illustrate the program. Details like including sex in the species are ignored. Working out the general solution for the stationary distribution equation and studying its stability properties is under progress and the reader is urged to consult the source (Gueron & Levin, 1995) for more details.

2.1 The model

We consider a continuous model where a population of total size P is distributed into groups of size x ($x \in [0, \infty)$). Let $f(x,t)$ denote the density distribution of groups of size x at time t. Assuming that all of the moments (about 0) of $f(x,t)$ exist ($f(x,t)$ is constrained to have an exponentially decaying tail), we denote

$$M_n(t) = \int_0^\infty z^n f(z,t)dz = \text{the } n^{\text{th}} \text{ moment about 0.} \tag{1}$$

Further,

$$G(t) = M_0(t) = \int_0^\infty f(z,t)dz = \text{Total number of groups} \tag{2}$$

131

$$P(t) = M_1(t) = \int_0^\infty z f(z,t) dz \quad = \quad \text{Total population size} \tag{3}$$

For now, we restrict attention to fixed population size and so $P(t) = P$.

In the following, we discuss the stationary distributions ($f(x)$) and their associated moments (M_n). Their time-independence is distinguished by omittance of explicit time-dependence.

The model rests on two density-dependent functions,

$\psi(x,y)$ = probability per unit time for two groups of sizes x and y to merge.

$\phi(x,y)$ = probability per unit time for a group of size x to fragment and generate groups of size y and $(x-y)$ (ignoring breakage into multiple fragments).

Of course,

$$\phi(x,y) = \phi(x, x-y) \quad \text{(by definition)} \tag{4}$$

$$x < y \Rightarrow \phi(x,y) \equiv 0 \quad \text{(fragments are smaller than the "mother" group)} \tag{5}$$

$$\psi(x,z) = \psi(z,x) \quad \text{(by symmetry)} \tag{6}$$

We also introduce

$p(x)$ = probability per unit time of a group of size x to break into two fragments,

$Q(x,y)$ = probability that, given that a group of size x breaks into two fragments, the fragments are of size y and $x - y$.

Clearly,

$$\phi(x,y) = p(x)Q(x,y) \tag{7}$$

and $p(x)$ is simply half (because of double counting) the integral of ϕ over all possible group sizes smaller than x i.e.,

$$p(x) = \frac{1}{2} \int_0^x \phi(x,z) dz = \frac{1}{2} p(x) \int_0^x Q(x,z) dz \tag{8}$$

so that, provided $p(x) \neq 0$,

$$\int_0^x Q(x,z) dz = 2. \tag{9}$$

2.1.1 Evolution equation for the group size distribution

The time evolution of the density distribution of groups is the result of the balance between fission and fusion :

$$\begin{aligned}
\frac{\partial f(x,t)}{\partial t} = &\; -p(x)f(x,t) - \int_0^\infty f(x,t)f(z,t)\psi(x,z)dz \\
&+ \frac{1}{2} \int_0^x f(y,t)f(x-y,t)\psi(y,x-y)dy + \int_x^\infty f(y,t)\phi(y,x)dy
\end{aligned} \tag{10}$$

132

where the four terms reflect respectively fragmentation of groups of size x, amalgamation of groups of size x with other groups, generation of groups of size x by merging of two smaller groups and fragmentation of larger groups into fragments, one of which is of size x.

2.1.2 Equation for the stationary distribution

By setting the time-derivative in the above equation to zero, we obtain the equation for the stationary distribution, $f(x)$:

$$-p(x)f(x)-f(x)\int_0^\infty f(z)\psi(x,z)dz+\frac{1}{2}\int_0^x f(y)f(x-y)\psi(y,x-y)dy+\int_x^\infty f(y)\phi(y,x)dy = 0 \tag{11}$$

We are interested in the structure of the distribution — to see if there is a peak corresponding to the "most frequent group size" which appears endogenously from the social dynamics. It would be instructive to study how and when this peak appears, what determines its position and width, whether multiple peaks are admissible and so forth. More generally, we would like to examine the existence, uniqueness and stability properties of the stationary distribution.

2.2 Simplified scenario

In order to make the model more tractable, we assume that $Q(x,y)$ is independent of y i.e., the conditional distribution for fragments is uniform. Hence, $Q(x,y) = \frac{2}{x}$, implying $\phi(x,y) = \frac{2}{x}p(x)$. Further, we assume that the symmetric function $\psi(x,y)$ is separable, so that $\psi(x,y) = \alpha a(x)a(y)$, where $a(x)$ is some nonnegative function. Also, we restrict attention to the special form for $p(x)$ viz., $p(x) = \beta x a(x)$ which implies that $\phi(x,y) = 2\beta a(x)$. α and β are some positive constants.

With this, the stationary distribution equation becomes

$$-\beta x a(x)f(x) - \alpha a(x)f(x)\int_0^\infty a(z)f(z)dz$$

$$+\frac{1}{2}\alpha\int_0^x a(y)f(y)a(x-y)f(x-y)dy + 2\beta\int_x^\infty a(y)f(y)dy \;=\; 0 \tag{12}$$

At this point, any specific case of the model is defined by the choice of the function $a(x)$ and the parameters α and β.

2.2.1 Exact solution

In the above simplified equation for the stationary distribution, notice that $a(x)$ and $f(x)$ always appear together with the same arguments and so, if we define the auxiliary function

133

$g(x) \equiv a(x)f(x)$, the equation reduces to

$$-\beta x g(x) - \alpha g(x) \int_0^\infty g(z)dz + \frac{1}{2}\alpha \int_0^x g(y)g(x-y)dy + 2\beta \int_x^\infty g(y)dy = 0 \qquad (13)$$

which incidentally is the equation for $f(x)$ if $a(x) = 1$.

The solution by the method of Laplace transform is

$$g(x) = 2\frac{\beta}{\alpha}\exp\left[-\left(2\frac{\beta}{\alpha P}\right)^{\frac{1}{2}}x\right] \qquad (14)$$

Returning to the general case we write the stationary solution

$$f(x) = 2\frac{\beta}{\alpha}\frac{1}{a(x)}e^{-\lambda x} \qquad (15)$$

where the value of λ is determined by the equation

$$P = 2\frac{\beta}{\alpha}\int_0^z \frac{z}{a(z)}e^{-\lambda z}dz \qquad (16)$$

This result is valid for any positive $a(x)$ with algebraic asymptotic behavior at $x \to \infty$ (to ensure all moments exist). Singularity occurs when $a(0) = 0$ (e.g., for $a(x) = x$), although the integral may still converge. Consequently, to ensure that the population (P) remains finite, we must restrict the choice $a(x)$ to satisfy

$$\lim_{x\to 0}\frac{a(x)}{x^2} > 0 \qquad (17)$$

This explicit solution proves the existence and uniqueness for the stationary distribution.

From equation (15), it is easy to see that for every choice of nondecreasing function $a(x)$, the resulting stationary distribution is a monotonically decreasing function of x. On the other hand, we conclude that any choice of a decreasing function $a(x)$ yields a group size stationary distribution that has at least one local maximum. Below, we look at some concrete examples of $a(x)$ to bring these points out.

2.2.2 Three typical examples

We focus our attention on three typical examples, where the function $a(x)$ is constant, decreasing, or increasing. Figure 1 shows the stationary distributions in these three cases.

Case 1 : $a(x) = 1$

Here, we have $\psi(x,y) = \alpha, p(x) = \beta x$, and $\phi(x,y) = 2\beta$. In this case, group fusion is independent of group size; that is the probability for a group of size x to merge with a group
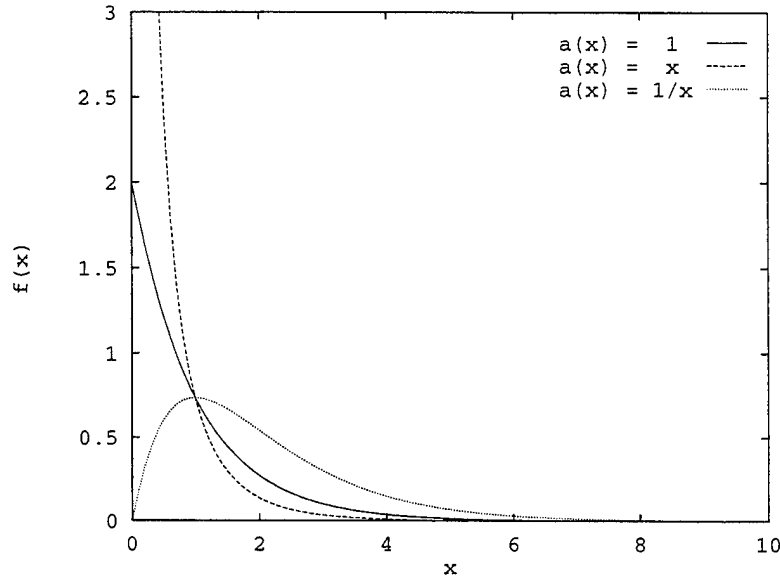
Figure 1: Stationary distributions for $a(x) = 1$ (exponential distribution), $a(x) = x$ (singular distribution), and $a(x) = 1/x$ (distribution with local maximum). $\alpha = \beta = \lambda = 1$.

of size y is constant (independent of x and y). Furthermore, group fission is proportional to group size.

The stationary distribution is an exponential distribution of group sizes, in which the most frequent groups are the "small ones".

Case 2 : $a(x) = x$

Here, we have $\psi(x, y) = \alpha x y$, $\phi(x, y) = 2\beta x$, and $p(x) = \beta x^2$. This accounts for grouping behavior in which large groups are more "attractive" or likely to collide and merge with others than are smaller ones, in proportion to their size. Furthermore, larger and larger groups become increasingly unstable, so that the fragmentation rate increases nonlinearly with group size.

The stationary distribution is singular at $x = 0$, i.e., we have small groups piling up near $x = 0$ and larger groups becoming less abundant.

Case 3 : $a(x) = 1/x$

Here, we have $\psi(x, y) = \alpha/xy$, $\phi(x, y) = 2\beta/x$, and $p(x) = \beta$. Thus, groups break at a rate independent of their size, for example due to external disturbances that affect all groups uniformly. Furthermore, the form of $\psi(x, y)$ suggests that solitary individuals, or individuals in small groups, do not remain in such groups for long. Possibly, due to the fact that it is disadvantageous to be in small groups, small groups merge at high rates to form larger ones.

The stationary distribution has a local maximum corresponding to the modal group size.

135

2.3 Summary

The problem of what determines group size distribution in nature is a fundamental one, carrying with it both ecological and evolutionary significance. Here, we see that a population can display a modal group size without having this quantity built explicitly into the dynamics. As individual behaviors change in response to cues, so too will the distribution of groups formed.

3 The timing of life history events

This section, drawing upon Iwasa & Levin (1995), investigates evolutionarily stable seasonal timing of life history events (emergence, germination, breeding or migration to find new sites). In each of these cases, the "early birds" have the greatest potential for high payoffs, and for securing favorable habitat, but may run the greatest risks of uncertain conditions. We suppose that following reproduction or germination there is a risk of disturbance, for example due to frost, storm, or predation and that such a disturbance event would kill all offspring already produced. If the last date of disturbance were known beforehand, the evolutionary tradeoff would be easily resolved : all individuals would germinate or breed just after the last distubance event. However, uncertainty exists regarding dates of disturbance. Hence, evolution will select for a compromise, balancing the tradeoffs between the higher risks and higher benefits of early germination.

Let x be the last date of disturbance, and $f(x)$ the probability distribution of x. We assume that $f(x)$ has a single peak. The latest possible date of disturbance is denoted by t_f. An example of such a distribution is

$$
f(x) = \begin{cases} C\, x^{c_1-1}\,(t_f - x)^{c_2-1}, & \text{for } 0 < x < t_f \\ \\ 0, & \text{otherwise} \end{cases} \tag{18}
$$

The evolutionarily stable strategy depends upon the details of interaction of environment with the population. Some of these are :

- **Spatial scale of disturbance** Disturbances may have a finite spatial extent and hence only affect individuals in a small area or may affect the entire population. We refer to the area affected by a single disturbance as a "patch".

- **Mode of competition** Fitness of individuals may be determined by time allowed for growth or by whether they managed to occupy nest sites.

- **Mode of population regulation** The dynamics of who wins and who loses in the game of life can occur over a small area or alternatively, in the presence of migration, over a large area.

3.1 Rank-independent fitness

First, we consider a scenario with

- **Spatial scale of disturbance** The environment consists of a number of patches, each having a different last date of disturbance. The distribution of number of patches having various last dates follows $f(x)$ and is constant over generations.

- **Mode of competition** Let $\psi(y \mid x)$ be the breeding success of an individual that begins breeding on day y given the last date of disturbance is x. The breeding success, provided individuals are not killed by the disturbance, decreases with the laying date y. A possible form for $\psi(y \mid x)$ is

$$
\psi(y \mid x) = \begin{cases} e^{-by}, & \text{if } y > x \\ \\ 0, & \text{if } y < x \end{cases}
\tag{19}
$$

- **Mode of population regulation** There is ample mixing and population regulation takes place effectively over the entire population.

Let $\phi(y)$ represent the expected fitness of individuals that breed at time y. This is simply the fitness given that a disturbance did not occur i.e.,

$$
\begin{aligned}
\phi(y) &= \int_0^{t_f} \psi(y \mid x) f(x)\, dx \\
&= e^{-by} \int_0^y f(x)\, dx
\end{aligned}
\tag{20}
$$

Because $\phi(y)$ is a non-negative continuous function on $[0, t_f]$, it generally attains its maximum at a single date y^*. This suggests that all the individuals should evolve to start breeding synchronously on the same date. Some of them are killed because the disturbance in their patch comes later than y^*, but reproduction in other patches in which the disturbance ends before y^* will produce the next generation. The population on the whole is quite stable in recruitment because the fraction of disturbed and nondisturbed patches does not change between generations.

3.2 Rank-dependent fitness

Next, we consider the following scenario :

- **Spatial scale of disturbance** There is only one patch i.e., the disturbance affects the whole area.

- **Mode of competition** The relative advantage of different breeding dates is given not by the absolute date but by the ranking within the local population. To illustrate this, suppose that a limited number (n) of nesting sites are available, and that only those who start breeding in one of these sites are able to breed. Then, to achieve some reproductive success, one must be included among the first n individuals that start breeding after the last disturbance date x.

- **Mode of population regulation** Success in acquiring a nest site is the method of population regulation.

Suppose that the reproductive success of an individual that starts breeding on day y is

$$
\psi(y \mid x) = \begin{cases} e^{-a \int_x^y g(z)\,dz}, & \text{if } y > x \\ \\ 0, & \text{if } y < x \end{cases} \tag{21}
$$

where a is the parameter reflecting the intensity of competition and $g(y)$ is the distribution of breeding dates. A very large value of a indicates very strong competition so that only those who start breeding just after the last disturbance can have reasonable reproductive success. In contrast, if a is zero, there is no advantage to start breeding earlier.

The fitness of an individual with breeding date y is

$$
\begin{aligned}
\phi(y) &= \int_0^{t_f} \psi(y \mid x)\, f(x)\, dx \\
&= \int_0^y e^{-a \int_x^y g(z)\,dz}\, f(x)\, dx
\end{aligned} \tag{22}
$$

The evolutionarily stable population will be comprised of a distribution of individuals, each with a different starting date. To see this, assume otherwise : if the population were composed only of individuals starting on a single "best date" y^*, then a mutant that started breeding just before y^* would have a large advantage, enjoying the best growing success in the population at the cost of a very small increase of risk. Hence, a distribution of breeding dates is the only evolutionarily stable strategy, which we denote by $g^*(y)$. Further, this distribution will ensure that all breeding individuals have the same fitness (denoted by λ). This is true by definition because if there were a date with better or worse fitness, then the distribution would be unstable. This implies that

$$
\begin{aligned}
\phi(y) &= \lambda, \quad \forall y \text{ such that } g^*(y) > 0 \\
\phi(y) &\le \lambda, \quad \forall y \text{ such that } g^*(y) = 0.
\end{aligned} \tag{23}
$$

It may be shown that

$$
g^*(y) = \begin{cases} f(y)/\int_{x_c}^{t_f} f(x)\,dx, & \text{for } y > x_c \\ \\ 0, & \text{for } y < x_c \end{cases} \tag{24}
$$

138

where the critical date x_c is given by

$$\int_{x_c}^{t_f} f(x)\, dx = \frac{a}{1+a} \qquad (25)$$

Figure 2 illustrates a couple of examples of this strategy. Equation (25) indicates that x_c becomes very early (close to zero) if competition is intense (large a); in this case, $g^*(y)$ approaches $f(y)$. In contrast, if the competitive advantage is small (small a), x_c approaches t_f and breeding should occur only after the disturbances have almost surely passed.
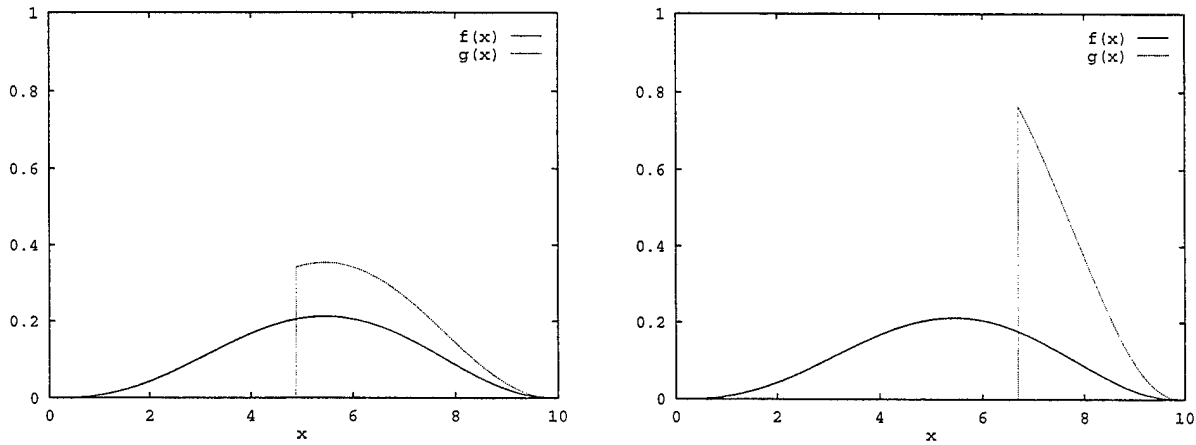

Figure 2: The evolutionarily stable distribution of starting date for rank-dependent fitness scenario. $f(x) = 3 \times 10^{-5}\ x^3\ (10 - x)^{2.5}$. (a) competitive advantage of starting breeding early is large and truncation thus occurs relatively early. (b) competitive advantage is small and thus the truncation date is late.

3.3   Locally regulated population

Here, we have the same scenario as in the rank-independent fitness case except for the mode of population regulation — there is no mixing prior to population regulation and so population is regulated on a local basis within each patch or subhabitat. This is termed "soft" selection in contrast to the "hard" selection occurring in the rank-independent case.

The reproductive success of an individual with starting date y is proportional to that given in Eq. (19), but is normalized within each local population :

$$\psi(y \mid x) = \begin{cases} e^{-by} / \int_x^{t_f} g(z)\, e^{-bz}\, dz, & \text{if } y > x \\[2mm] 0, & \text{if } y < x \end{cases} \qquad (26)$$

The reproductive success of an individual with breeding date y is the average of ψ over the

different local populations included in the whole population :

$$\phi(y) = \int_0^y \frac{e^{-by}}{\int_x^{t_f} g(z)\, e^{-bz}\, dz}\, f(x)\, dx \tag{27}$$

The evolutionarily stable distribution satisfies the same conditions (given by Eq. (23)) as in the rank-dependent case. It can be shown that the distribution satisfying these conditions is again a truncated one, where no breeding occurs before a critical date x_c but some oviposition occurs every day after x_c. x_c satisfies

$$f(x_c) = b \int_0^{x_c} f(y)\, dy \tag{28}$$

and the evolutionarily stable distribution of oviposition dates is

$$g^*(y) = \begin{cases} \frac{-e^{by}}{b\lambda} \frac{d}{dy}\left(f(y)\, e^{-by} \right), & \text{if } y > x_c \\[2ex] 0, & \text{if } y < x_c \end{cases} \tag{29}$$

where λ is determined by

$$\lambda = \frac{f(x_c)}{b} + \int_{x_c}^{t_f} f(y)\, dy \tag{30}$$

Figure 3 illustrates a couple of examples of this strategy. The stable solution $g^*(y)$ is determined by the environmental disturbance function $f(x)$ and growth advantage b. x_c decreases with b — if b is strong, then the critical date x_c becomes early and $g^*(y)$ approaches $f(y)$ but has a peak later than the peak of $f(y)$. On the other hand, if b is weak, x_c approaches t_f and most breeding starts on the date all possible disturbance ends.

3.4   Synchronized disturbance

The scenario addressed here is similar to the rank-independent fitness case with the difference that the disturbances are synchronized over the whole population i.e., there is effectively only one patch to consider. The arithmetic average of reproductive success is no longer the correct criterion for measuring evolutionary advantage; instead, we must use the geometric average because the function maximized is the average of logarithmic reproductive success (assuming the population has no overlapping generations). Therefore, we have

$$\phi(y) = \int_0^{t_f} \log\left[\psi(y \mid x) \right] f(x)\, dx \tag{31}$$

The fitness $\psi(y \mid x)$ is given as before by Eq. (19). The evolutionarily stable strategy is a distribution and can be specified by a probability distribution $p(y)$. The logarithmic fitness is given by a functional of $p(y)$:

$$\phi(p(\cdot)) = \int_0^{t_f} \log\left[\int_x^{t_f} e^{-by} p(y)\, dy \right] f(x)\, dx \tag{32}$$
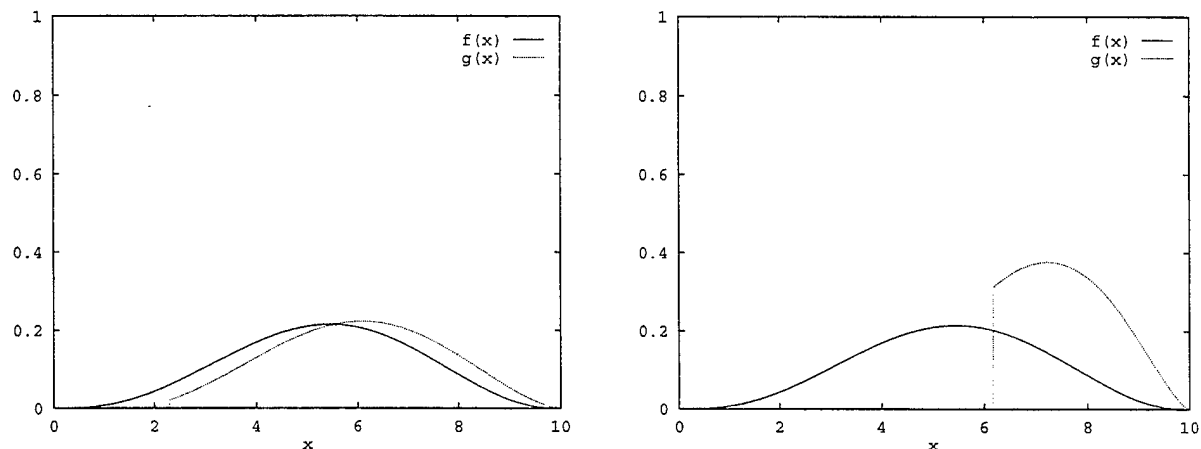
140

Figure 3: The evolutionarily stable distribution of starting date for locally regulated population scenario. $f(x) = 3 \times 10^{-5} \ x^3 \ (10 - x)^{2.5}$. (a) growth advantage is large ($b = 1.5$) and the truncation occurs relatively early. (b) growth advantage is small ($b = 0.3$) and the truncation date is late. If $g(y)$ is regarded as a probability distribution for a mixed strategy, the same graph gives the optimal mixed strategy for the synchronized disturbance case.

The evolutionarily stable strategy is the one that maximizes the above functional under the constraint

$$\int_0^{t_f} p(y)dy = 1 \text{ and } p(y) \geq 0 \tag{33}$$

This constrained optimization problem can be analyzed via Langrange's multipliers. It has the same mathematical structure as Eq. (23) and hence the optimal strategy is the same as obtained for the locally regulated population case. However, there is a crucial difference — in the case of locally regulated population both a **mixed strategy** (breeding dates following a distribution) and a mixture of **pure strategies** (all members of subpopulation breeding synchronously) achieve an evolutionary equilibrium whereas in the present case of bet-hedging, a single mixed strategy is optimal but a mixture of pure strategies is unstable.

3.5   Discussion

To summarize, we see that in the rank-independent case the optimal strategy calls for synchronous breeding-start. In other situations, asynchronous breeding evolves — either as a mixture of individuals having pure strategies or as a population choosing the breeding date according to a probability distribution.

Another way to slice the problem is to consider the effects of $S_{\text{regulation}}$, the spatial scale over which population regulation occurs versus $S_{\text{disturbance}}$, the patch-size. If $S_{\text{regulation}} \gg S_{\text{disturbance}}$ a pure strategy is expected to evolve. If, on the other hand, $S_{\text{regulation}} \leq S_{\text{disturbance}}$ a mixed strategy or a mixture of pure strategies is expected to evolve.

141

4 Self-organization of front patterns in large wildebeest herds

Aerial photographs of migrating wildebeest herds reveal striking distributional patterns. Some share the same characteristic zigzagged irregular wavy front, with a typical wavelength large with respect to the body-length (e.g., see figure 4). These patterns vary over scales that are much larger than the individual's perceptual range, and thus cannot be explained simply as random fluctuations on uniformity. Instead, the patterns seem to be self-organizing, building and reinforcing on inhomogeneities. Furthermore, since the individual is only aware of its immediate surroundings, broad-range patterns must be explained in terms of local decisions. This section drawing upon its source (Gueron & Levin, 1993) proposes such a model that exhibits spontaneous generation of long-range patterns similar to observed herd front patterns and provides a mechanism for it.

Figure 4: An aerial photograph of a large wildebeest herd, revealing large-scale patterns. (reprinted from Sinclair, A. R. E. (1977) by courtesy of The University of Chicago Press.)

4.1 The model

4.1.1 Tracking "leaders" is sufficient

Modeling of small herds (Gueron et. al., unpublished data) revealed that the herd "leaders" (i.e., the individuals at the front, with respect to the direction of propagation) belong to a distinct group, namely those that (for whatever reason) inherently have a faster walking pace (speeders). Without any global plan (or knowledge of the global distribution of the herd), speeders simply advance, overtaking slower individuals, and finally generate a leading band that consists of only speedy individuals. The propagation direction is controlled mainly by these leaders. This allows for some simplifications in modeling wildebeest herds at large scales. Since the pattern development we are addressing is defined by leaders, we regard *only* the "leading band" for analyzing this phenomena. The rest of the herd (i.e., the trailing layers), which is not modeled explicitly, lies behind the front layer and "fills in the gaps" as they develop.

The front layer is modeled as a curve $y(x,t)$, $x \in (-\infty, \infty)$, $t \geq 0$ evolving in both space and time. In this model we assume that the speedy band has "perfect directionality". Hence, changes in the velocity are reflected only by changes in its magnitude (while the direction remains constant). Without loss of generality, this direction is chosen to be the Y-axis.

4.1.2 Speed adjustment

Individuals are considered to have an inherent preferred speed at which they walk when they do not interact with their (left and right) neighbors. When they do interact, they adjust (increase or decrease) their speeds according to an intrinsic behavior pattern. If for instance, individuals trailing relative to their neighbors speed up and those leading slow down, the pattern at large time would correspond to the leading edge moving as a straight front. If on the other hand, individuals trailing slow down and leaders speed up, any inhomogeneity will be reinforced. Provided that these behaviors reverse when the gap increases too much, the instability will saturate and keep the herd from segregating. Segregation will occur if encouraged speeders go on speeding and never stop to "wait" for the trailers to catch up.

The interaction among the herd members is modeled as a function of the relative position with respect to the neighbors. The relative position as defined by

$$\Delta(y(x,t)) = \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} y(s,t)ds - y(x,t), \tag{34}$$

which compares the location of an individual located at $y(x,t)$ with the average location of the neighborhood (ranging δ to each side) surrounding it. The (positive) parameter δ determines how local the interaction with the neighbors is. Here, δ is small compared with the characteristic wavelength of the front pattern, thus accounting for a local influence range.

The velocity $(v = \dot{y})$ is modeled as a function of Δ. A straightforward way to model an intrinsic velocity $(v_0(t))$ and the ability of an individual to adjust its speed as a result of interaction with its neighbors is

$$\dot{y} = v_0(t) + F(\Delta(y)) \tag{35}$$

for some (differentiable) function F reflecting the speed adjustment that depends on the interaction.

4.2 Stability analysis of traveling front solutions

The natural question to ask at this point is whether there are simple steady-state solutions that produce "wavy" front patterns and, in particular, vary over scales that are larger than $O(\delta)$. If we restrict attention to solutions of polynomial order (i.e., solutions that are bounded by $c|x|^n$ as $|x| \to \infty$, for some positive integer n and a real constant c), it can be shown that there are no bounded solutions on $x \in (-\infty, \infty)$ which yield the multiple peaks we seek. The only bounded solution of interest from a stability standpoint is the uniform traveling front (front moving as a straight edge). If $F'(0) = b \neq 0$, then the uniform traveling front is (linearly) stable for $b > 0$ and unstable for $b < 0$. Further, wavelength of the fastest growing mode when $b < 0$ is estimated to be

$$\mu_{\max} \sim \frac{4.493}{\delta}. \tag{36}$$

This estimate gives a rough approximation to the characteristic wavelength observed in the nonlinear problem.

4.3 Numerical results

For numerical computations, Eq. (35) is transformed to the discrete form

$$\dot{y}_i = v_0(t) + F(\Delta(y_i)), \quad i = 1 \ldots n \tag{37}$$

where Δ is written as

$$\Delta(y_i(x, t)) = \frac{1}{2p + 1} \sum_{j=i-p}^{j=i+p} y_j - y_i. \tag{38}$$

The continuous (infinite) region $x \in (-\infty, \infty)$ is replaced with the finite set of samples at $i = 1 \ldots n$ with periodic boundary conditions. p is the number of neighbors on each side of the animal with which it interacts. Equations (37) and (38) form a system of n coupled ordinary differential equations for the variables y_i representing the herd's front layer. These equations are written in dimensionless form. We define our length and time units in such a

144

way that the intrinsic velocity, v_0, equals 1. In the simulations, a dimensionless time step of 0.1 is used.

F, the response to displacement from a uniform position, satisfies $F(0) = 0$. It should also be bounded. Since it represents the velocity change, corresponding to a given displacement, its magnitude should not exceed v_0 (to avoid negative velocities, or exceeding twice the intrinsic velocity). F should have one positive and one negative simple root to provide for reversing the behavior trends when the gap increases too much. In the simulations, the following function that satisfies the above requirements is used,

$$F(\Delta) = \pm \sin(\Delta) \qquad (39)$$

where the sign $\pm$ is chosen as $+$ to demonstrate the stable cases (i.e., where the uniform traveling front is a stable solution) and $-$ for the unstable cases (where inhomogeneities in the uniform traveling solution get amplified).

4.3.1 Stable case — straight front

For the stable case, the response to lagging behind is speeding up, and the response to being ahead is slowing down. As implied by the stability analysis, the uniform traveling front is a stable solution for such cases. It is expected to be an "attractor" in the sense that any initial conditions will eventually lead to a uniform front after a long enough time has elapsed. Figure 5 presents the results of a run where the initial pattern decays to a uniform front.
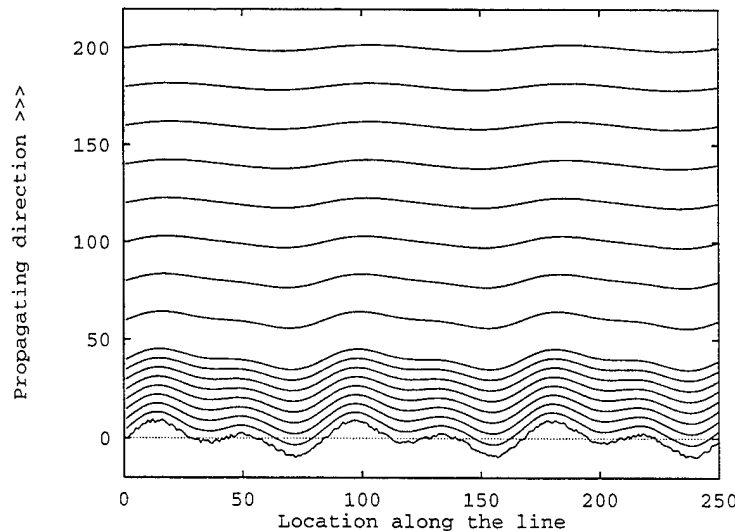

Figure 5: The stable case. Traces of $y_i(t)$ versus i for 17 different time steps from $t = 0, \ldots, 200$. The initial disturbances get smoothed out to form a straight front.

4.3.2 Unstable case — irregular wavy front

In the unstable cases, the response to lagging behind is slowing down (until the gap reaches the saturation level where the behavior is reversed). On the other hand, being ahead is "encouraged" by speeding up (until the gap reaches the saturation level where the behavior is reversed). For such cases the uniform front is an unstable solution, and small perturbations are expected to grow and to generate some irregular patterns. Figure 6 shows the runs with three different values for p. We see that, indeed, we get "wavy" fronts where the typical wavelength is on a larger scale than the typical interaction range. Thus, local rules have produced long-range effects. It is also seen from the simulations that after bifurcating from the uniform arrangement, these solutions reach a "semi-steady state" i.e., although the fronts keep changing in time, the changes are on a small scale and the global pattern seems persistent.
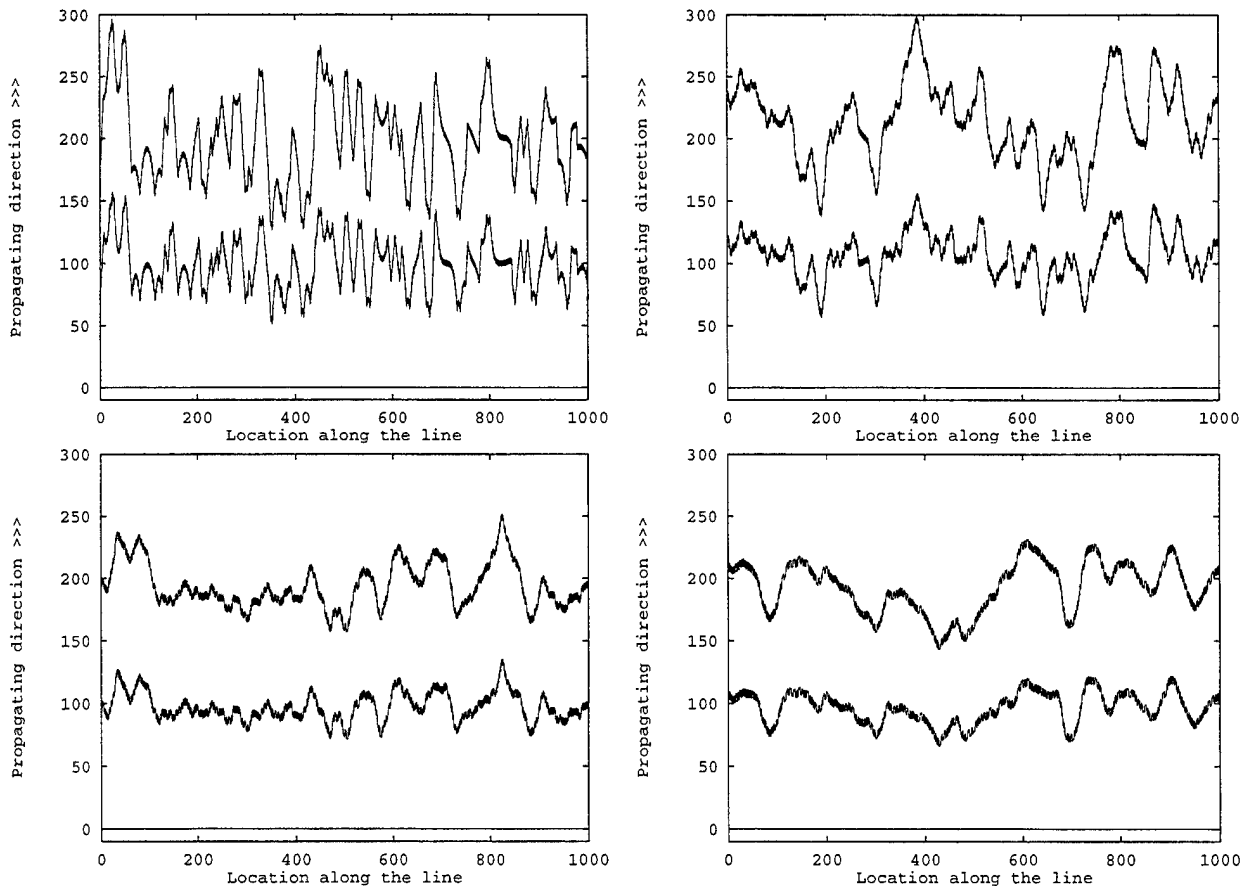

Figure 6: The unstable case. Traces of $y_i(t = 0, 100, 200)$ versus i for p = 1 (upper left), p = 2 (upper right), p = 3 (bottom left), and p = 4 (bottom right). Smooth initial conditions produce irregular wavy patterns. Increasing the local interaction neighborhood increases the wavelength of the large-scale pattern.

4.3.3 Effect of size of interaction neighborhood

Comparing plots in figure 6 reveals that the ratio of the observed characteristic wavelengths is roughly as predicted from eqn. (36). Further, computer experiments also suggest that as the interaction becomes more local, the patterns become more erratic and hence the front becomes longer. In the limit that the interaction neighborhood becomes very large, the front becomes very flat. Of great interest is the fact that the patterns fit the experimental observations better as the neighborhood of interaction becomes more local.

5 The importance of being discrete and spatial

One of the fundamental issues in the modeling of any system is the choice of level of detail. Figuring out which details at one level are important to the determination of phenomena at other levels, and which can be ignored lies at the heart of understanding mechanisms. In this section, based on Durrett & Levin (1994), we consider and compare four approaches to modeling the dynamics of spatially distributed systems : 1) mean field approaches (described by ODE's) in which every individual is considered to have equal probability of interacting with every other individual; 2) patch models that group discrete individuals into patches without additional spatial structure; 3) reaction-diffusion equations, in which infinitesimal individuals are distributed in space; and 4) interacting particle systems, in which individuals are discrete and space is treated explicitly. Each model represents different assumptions about the interactions at the microscopic/individual scale. The predictions about macroscopic quantities are studied in three separate biological scenarios.

5.1 Four approaches to modeling species interactions

We pose a system in which there are two types of individuals, whose interaction is described by a game matrix :

$$
\begin{array}{cc}
 & \begin{array}{cc} H & D \end{array} \\
\begin{array}{c} H \\ D \end{array} & \begin{array}{cc} a & b \\ c & d \end{array}
\end{array}
\tag{40}
$$

Here, H denotes a hawk and D a dove. This nomenclature is applicable in scenario 3 below, but more generally it is used to refer to strategies employed by two sub-populations. To explain the game matrix : b is, for example, the payoff to a hawk when interacting with a dove. When the population consists of a fraction p of hawks and $1 - p$ of doves then the payoff for an individual hawk is $ap + b(1 - p)$ which is interpreted as the net birth/death of hawks.

In addition to the net growth rate, we assume that there is a density dependent death rate per individual that is proportional to the total density.

147

Below, we describe the four approaches taken. See figure 7 for relationships between the approaches.
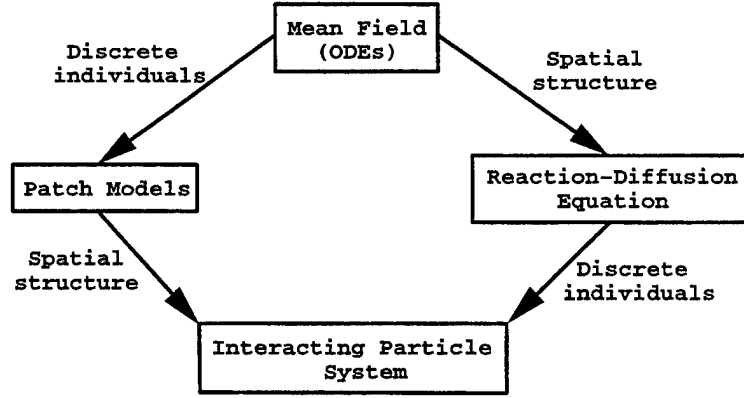

Figure 7: Relationship of the four approaches studied.

In our first two models, we assume that the populations are large enough to ignore the discreteness of individuals. Under the continuum hypothesis, we speak of density of individuals as a real-valued quantity and interpret rates as terms in the differential equation for the density.

5.1.1 Mean field model

If we ignore all spatial structure, and lump all the individuals together so that every individual has an equal probability of encountering every other individual, then we arrive at the following dynamical system for the densities of hawks (u) and doves (v) :

$$
\begin{aligned}
\frac{du}{dt} &= u\left\{a\frac{u}{u+v} + b\frac{v}{u+v} - \kappa(u+v)\right\} \\
\frac{dv}{dt} &= v\left\{c\frac{u}{u+v} + d\frac{v}{u+v} - \kappa(u+v)\right\}
\end{aligned}
\tag{41}
$$

5.1.2 Reaction-diffusion system

The traditional (but occasionally incorrect) way to turn a dynamical system into a reaction-diffusion system is to simply add diffusion terms to the two equations :

$$
\begin{aligned}
\frac{du}{dt} &= \Delta u + u\left\{a\frac{u}{u+v} + b\frac{v}{u+v} - \kappa(u+v)\right\} \\
\frac{dv}{dt} &= \Delta v + v\left\{c\frac{u}{u+v} + d\frac{v}{u+v} - \kappa(u+v)\right\}
\end{aligned}
\tag{42}
$$

where $\Delta = \partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$. For simplicity we have assumed that the two diffusion constants are the same and we have scaled time to make them equal to 1. When different diffusion rates are considered, an additional parameter (the ratio of the diffusion constants) enters and the range of behaviors is enlarged.

148

In the next two models, individuals retain their nature as discrete units. Let S be the set of sites or "patches", and assume that the state at time t is given by two functions η_t and ζ_t from S to $\{0,1,...\}$. So, $\eta_t(x)$ and $\zeta_t(x)$ represent the number of hawks and doves in patch/site x at time t. We take time to be continuous, i.e., t can be any nonnegative real number; so the temporal evolution is described by specifying the rate at which things happen. We say something happens at rate r if the probability of an occurrence in a short amount of time h is $rh + o(h)$, where $o(h)$ denotes a quantity with $o(h)/h \to 0$ as $h \to 0$. When an event occurs at a constant rate r then the times t_i between successive occurrences have an exponential distribution with parameter r; that is, $P(t_i > t) = \exp(-rt)$. In the following two models, three types of events occur, viz., migration, deaths due to crowding, and the game step. The difference between the two models lies in their treatment of space.

5.1.3 Patch model

Subdivision of population into patches recognizes the importance of space at local scales but the collection of patches, S, has no spatial structure; thus, we choose $S = \{1, 2, ..., N\}$ where N is the number of patches.

Migration. Each individual changes its spatial location at rate μ, and when it moves, it moves to a randomly chosen patch.
Deaths due to crowding. Each individual at x at time t dies at rate $\kappa(\eta_t(x) + \zeta_t(x))$.
Game step. Let $p_t(x) = \eta_t(x)/(\eta_t(x) + \zeta_t(x))$ be the fraction of hawks at site x. Each hawk experiences a birth (or death) rate of $a p_t(x) + b(1 - p_t(x))$ while each dove experiences a birth (or death) rate of $c p_t(x) + d(1 - p_t(x))$.

5.1.4 Interacting particle system

Here, we identify patches with the two-dimensional integer lattice, i.e., $S = \mathbf{Z}^2$. The dynamics are formulated as follows :

Migration. Each individual changes its spatial location at rate μ, and when it moves, it moves to a randomly chosen neighbor of x.
Deaths due to crowding. Each individual at x at time t dies at rate $\kappa(\eta_t(x) + \zeta_t(x))$.
Game step. Let $\mathcal{N}$ be the interaction neighborhood for the model. We consider the following two choices for $\mathcal{N}$:

$$\mathcal{N}_1 = \{z \in \mathbf{Z}^2 : |z_1| + |z_2| \le 1\} \quad (0,0) \text{ and its nearest neighbors}$$
$$\mathcal{N}_2 = \{z \in \mathbf{Z}^2 : |z_1| \le 2, |z_2| \le 2\} \quad \text{a } 5 \times 5 \text{ square centered at } (0,0)$$

For any choice of $\mathcal{N}$ we define the number of hawks and doves in the interaction neighborhood of x at time t as

$$\hat{\eta}_t(x) = \sum_{z \in \mathcal{N}} \eta_t(x + z) \qquad \hat{\zeta}_t(x) = \sum_{z \in \mathcal{N}} \zeta_t(x + z)$$

and

$$p_t(x) = \hat{\eta}_t(x)/(\hat{\eta}_t(x) + \hat{\zeta}_t(x))$$

Like before, each hawk experiences a birth (or death) rate of $ap_t(x) + b(1 - p_t(x))$ while each dove experiences a birth (or death) rate of $cp_t(x) + d(1 - p_t(x))$.

A word about the initial conditions : in all the simulations with patch model and interacting particle system, the initial distribution of hawks and doves is specified by $[H_{\text{initial}}, D_{\text{initial}}]$ — this specifies a distribution where at each location, the number of hawks is a uniformly distributed integer in $[0, \ldots, H_{\text{initial}}]$ and likewise the number of doves is a uniformly distributed integer in $[0, \ldots, D_{\text{initial}}]$.

5.2   Scenario 1 — Stable equilibrium

$$
\begin{array}{ll}
a = 0.4 & b = 0.8 \\
c = 0.6 & d = 0.3
\end{array}
\qquad \kappa = 0.08
$$

This scenario represents an apparent symbiotic relationship because the fitness of each is enhanced by the presence of the other. This choice of a, b, c and d could also represent the sum of a constant growth rate and negative effects of competition :

$$
\begin{pmatrix} 0.4 & 0.8 \\ 0.6 & 0.3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix} + \begin{pmatrix} -0.6 & -0.2 \\ -1.4 & -1.7 \end{pmatrix}
$$

This would then correspond to the competition situation in which each species inhibits itself more than it inhibits the other, for which coexistence results owing to the existence of a globally stable internal equilibrium. This result is captured by all four of the models studied.

- **Mean field model** Figure 8 shows the behavior of the dynamical system in (u, v)-space. There is a unique equilibrium that is the limit starting from any initial state in which both species have positive density.

- **Reaction-diffusion system** Because the density-dependent dynamics promote coexistence locally, the reaction-diffusion system shows the same behavior globally as well.

- **Patch model** Figure 9 shows a simulation of the patch model with $N = 2500$ patches when $\mu = 1$ and we start from an initial condition $[4, 4]$. We observe that the number of hawks and doves rise from initial levels near 1.5 to equilibrate near 2.0 and 2.7 respectively; although there are fluctuations coming from the fact that there are only 2500 patches.
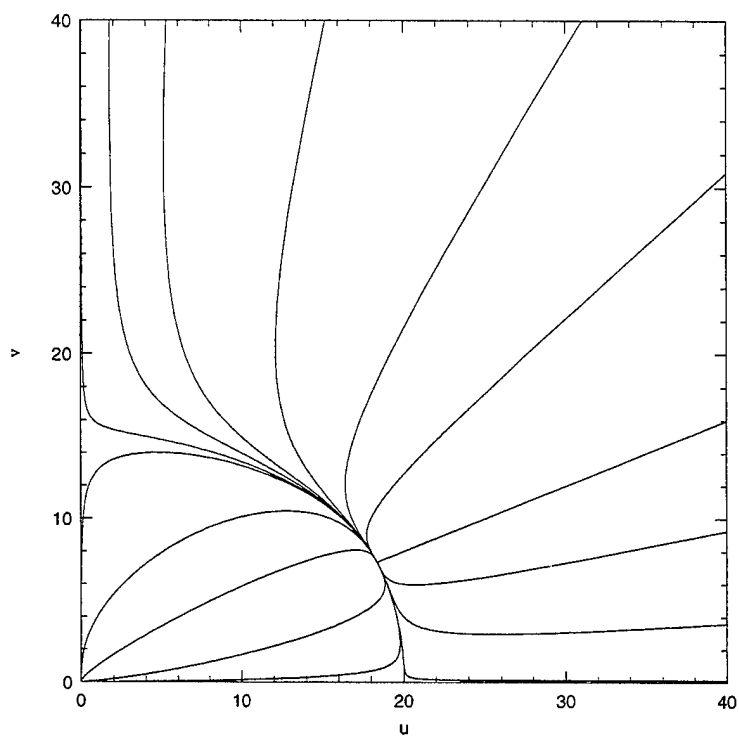
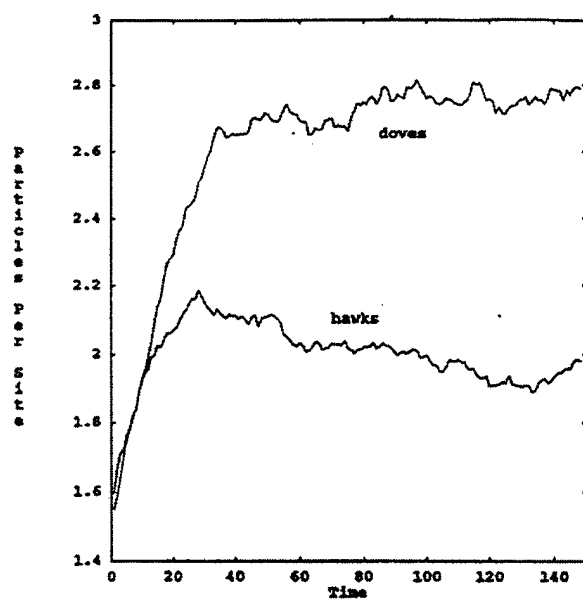Figure 8: Phase portrait for Scenario 1.

Figure 9: Average number of hawks and doves per site in Scenario 1 for a simulation of the patch model with 2500 sites.

- **Interacting particle system** Figure 10 shows a simulation on a 100×100 system with periodic boundary conditions with $\mu = 0.4$ starting from $[7.5, 7.5]$ initial condition. We observe that the densities of hawks and doves converge exponentially fast to their equilibrium values, although again there are some fluctuations in the densities due to the fact that there are only $10,000$ sites. It is suspected that this particular interacting particle system has a unique equilibrium state that is the limit starting from any initial state that is (a) translation invariant (spatially homogeneous) and (b) in which each species has a positive density.
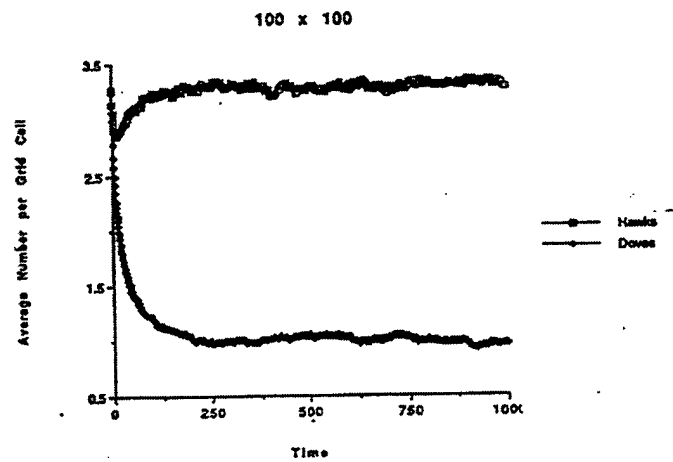

Figure 10: Average number of hawks and doves per site in Scenario 1 of the interacting particle system on a 100×100 lattice.

5.3 Scenario 2 — The importance of being spatial

$$
\begin{array}{ll}
a = 0.7 & b = 0.4 \\
c = 0.4 & d = 0.8
\end{array}
\qquad
\kappa = 0.08
$$

This may be thought of as two species competing for use of the same resource and in which competition each inhibits the other species more than it does itself. In this situation, the nonspatial models disagree with the ones that treat space explicitly.

- **Mean field model** Figure 11 shows that the dynamical system has two stable equilibria $(\bar{u}, 0)$ and $(0, \bar{v})$ on the axes, and their basins of attraction contain the whole positive quadrant except for a line through the origin containing the equilibrium (u^*, v^*), which is a saddle point. Here, initial densities determine the basin of attraction and hence which species wins out.
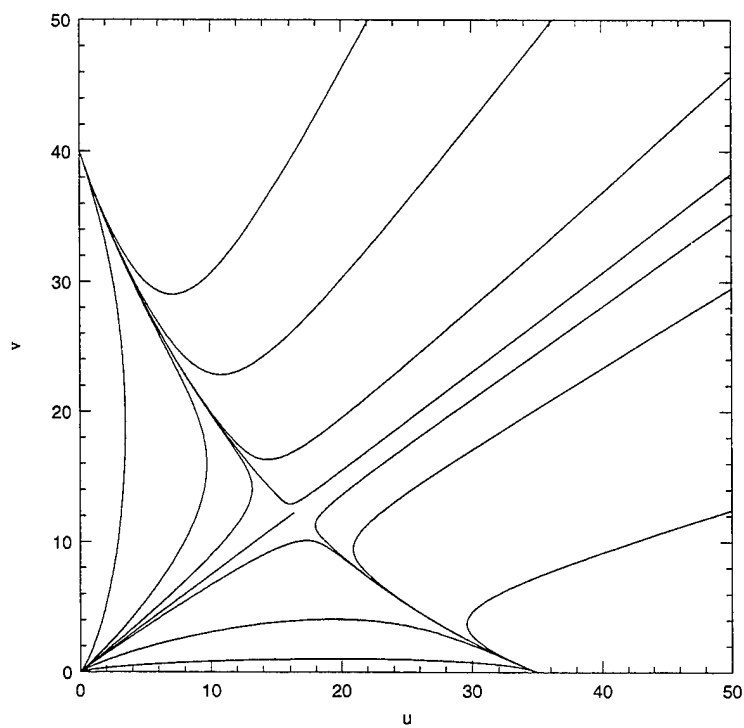
152

Figure 11: Phase portrait for Scenario 2.

- **Reaction-diffusion system** For "generic" initial conditions, $(u(x,t), v(x,t))$ will converge to $(0, \bar{v})$ uniformly on compact sets. To explain the eventual domination of doves, note that the reaction diffusion equation admits traveling wave solution $u(x,t) = U(x - \rho t), v(x,t) = V(x - \rho t)$, where U and V are monotone functions with $U(-\infty) = 0, U(\infty) = \bar{u}, V(-\infty) = \bar{v}, V(\infty) = 0$. Since $b = c$ and $d > a$, the velocity $\rho > 0$ indicating that an interface between a region of all doves and a region of all hawks moves in a direction that favors the doves.

- **Patch model** Figures 12 (a) and (b) show a simulation of the patch model with $\mu = 1$ for the $[5, 5]$ and $[7, 3]$ initial configurations respectively. In the first case the hawks dies out while in the second the doves dies out. Thus, the species that wins out depends on the initial conditions.
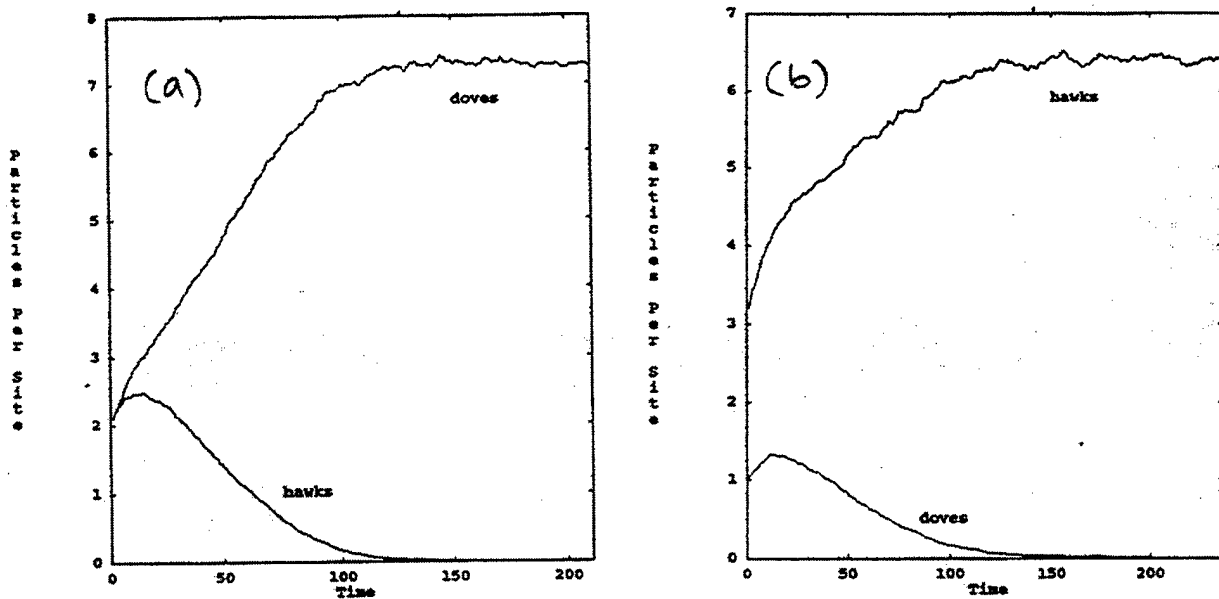

Figure 12: Average number of hawks and doves per site in Scenario 2 for a simulation of the patch model with 2500 sites. (a) $[5, 5]$ initial condition — hawks die out. (b) $[7, 3]$ initial condition — doves die out.

- **Interacting particle system** Figure 13 shows the behavior on a 100×100 system with periodic boundary conditions with $\mu = 1$ starting from $[10, 5]$ initial condition. As in the reaction-diffusion system, the doves win owing to the same mechanism — if a large enough pocket of doves (i.e., a region in space that contains mostly doves) forms then that region will grow linearly in radius and take over the system. On a small system this pocket may not have a chance to form, but if we consider the system on the infinite lattice $\mathbf{Z}^2$ and start with a translation invariant initial distribution with a positive density of doves, such a pocket will form with probability one and the hawks will die out.
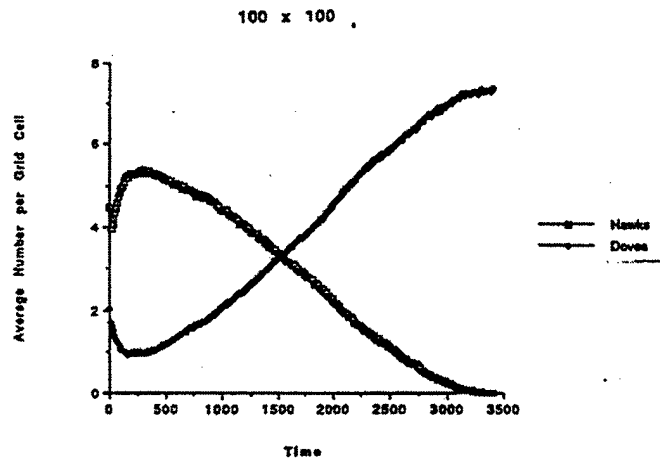
154

Figure 13: Average number of hawks and doves per site in Scenario 2 of the interacting particle system on a 100×100 lattice with $[10, 5]$ initial condition. The hawks die out.

5.4 Scenario 3 — The importance of being discrete

$$
\begin{array}{ll}
a = -0.6 & b = 0.9 \\
c = -0.9 & d = 0.7
\end{array}
\qquad \kappa = 0.02
$$

This corresponds to the case when the hawk strategy dominates the dove strategy — hawks always do better than the doves, but a population consisting purely of hawks dies out. This may be thought of as a classic fugitive species in a spatial mosaic or of two competitors sharing a resource along a successional gradient. In each case, the dominant species eliminates the weaker one locally but then dies out itself, and the vacant space is recolonized by the inferior species. Here, the models that explicitly take discreteness of individuals into account (patch model, interacting particle system) tell us that the hawks and doves can coexist in an equilibrium whereas the continuum models (mean field model, reaction-diffusion system) tell us that after large value of time, both populations die out.

- **Mean field model** Figure 14 shows that all trajectories starting with positive densities of hawks and doves end at $(0,0)$.

- **Reaction-diffusion system** It would seem conceivable to have a front of doves being chased by a front of hawks so that the area in front of the doves would not be colonized and the area behind the hawk population would be left empty. Numerical simulations conducted in a one-dimensional domain with no-flux boundary conditions confirm the existence of such a scenario (see figure 15). However, this solution becomes unstable when perturbed by even low-intensity noise in the hawk density and collapses to global
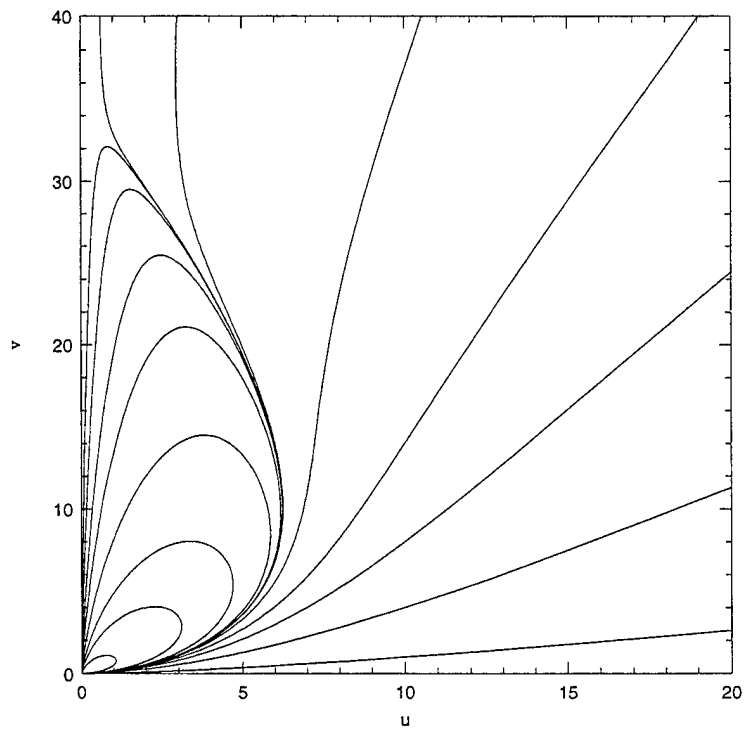
155

Figure 14: Phase portrait for Scenario 3.

156

extinction of both species. Hence, we conclude that since the local behavior tends towards extinction of both species, the reaction-diffusion system will collapse as well for "generic" initial conditions.
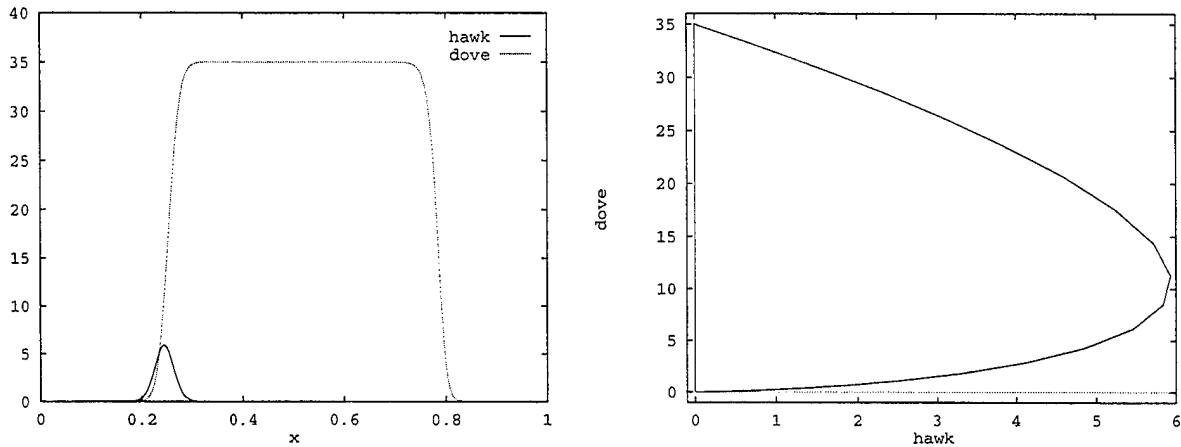

Figure 15: Solution of the reaction-diffusion system in Scenario 3 exhibiting a population of doves being chased by a population of hawks so that the area in front of the doves would not be colonized and the area behind the hawk population would be left empty. The direction of travel is to the right. The dove front moves faster than the hawk front and so this pattern does not become steady in some moving frame of reference. No flux boundary conditions were employed.

- **Patch model** Figure 16 shows a simulation of the patch model with $\mu = 1$ for the $[2, 8]$ initial condition. After an initial dramatic increase in the number of doves, the system settles down to an equilibrium where both hawk and dove densities are positive. The equilibrium is maintained as the doves recolonize empty spots where they are eventually attacked and taken over by hawks which eventually die out leaving the spot open for the cycle to start afresh.

- **Interacting particle system** A typical simulation of the interacting particle system begins with a period in which the hawk population grows faster than the dove population until the fraction of hawks is too large and both species start to die out. When the density gets low we have a few doves who are completely isolated and give birth at rate d. These doves start colonies that grow and would fill up the space to the doves' preferred equilibrium density, except for the fact that along the way they encounter a few hawks that managed to escape extinction. These hawks reproduce faster than the doves, the fraction of hawks grows, and the cycle begins again.

  Figure 17 shows the behavior on a 150×150 system with periodic boundary conditions with $\mu = 1$. We see that the hawk and dove densities oscillate around their equilibrium values. These oscillations grow as the system size is reduced and so can be called a *"finite-size effect"*. Essentially, the large oscillations are present in small sub-systems
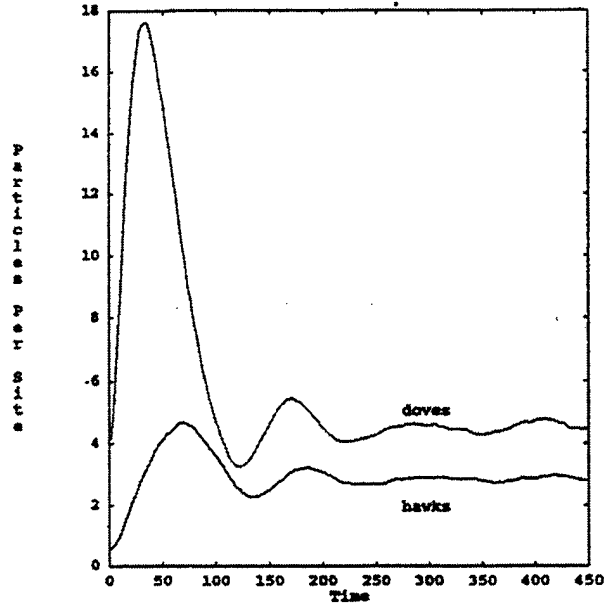
Figure 16: Average number of hawks and doves per site in Scenario 3 for a simulation of the patch model with 2500 sites and the [2,8] initial condition.

but they are not synchronized across sub-systems and hence cancel out yielding a constant average picture.

Also, the larger neighborhood over which averages are taken for the game step makes the local fraction of hawks less random and makes the system behave more like the dynamical system. To explain the last remark, consider our system on an $N \times N$ grid and suppose (i) each site uses an interaction neighborhood equal to the entire lattice and (ii) each particle experiences deaths at rate $\kappa(\hat{u}_t^N + \hat{v}_t^N)$ where $\hat{u}_t^N$ and $\hat{v}_t^N$ are the number of hawks and doves per site at time t. Then, it is easy to show that as $N \to \infty$, $\hat{u}_t^N$ and $\hat{v}_t^N$ converge to the solution of the differential equation (41).

We should also note that the equilibrium for the interacting particle system has spatial structure not possible in the patch model.

5.5 Partial differential equation limits of particle systems

It is possible to get alternative reaction-diffusion systems by applying the diffusion approximation to migration in the interacting particle system. If we consider particles performing independent random walks on $\mathbf{Z}^2$ then under suitable assumptions the joint distribution of the number of particles at any finite set of points will converge to independent Poisson random variables. Combining this observation with the fact that the migration occurs on a much faster time scale than the game interaction, sites near x at time t are independent and
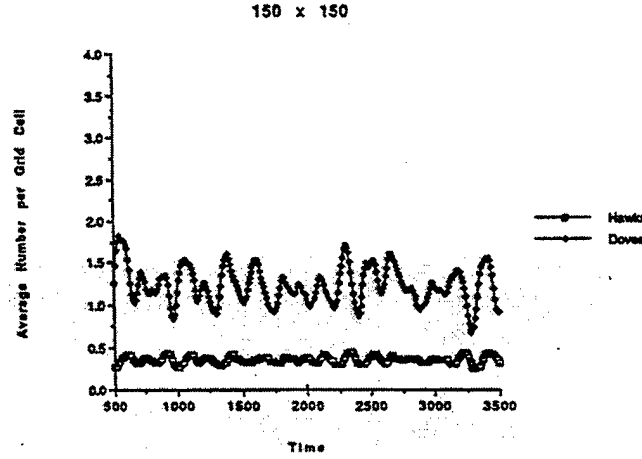
Figure 17: Average number of hawks and doves per site in Scenario 3 of the interacting particle system on a 150×150 lattice.

have a Poisson number of hawks and doves with means $u(x, t)$ and $v(x, t)$ which satisfy the "hydrodynamic equations" :

$$
\begin{aligned}
\frac{du}{dt} &= \Delta u + u \left\{ a \left(h + (1 - h)\frac{u}{u+v} \right) + b(1 - h)\frac{v}{u+v} - \kappa(1 + u + v) \right\} \\
\frac{dv}{dt} &= \Delta v + v \left\{ c(1 - h)\frac{u}{u+v} + d \left(h + (1 - h)\frac{v}{u+v} \right) - \kappa(1 + u + v) \right\}
\end{aligned}
\tag{43}
$$

where

$$
h = h(u, v) = \frac{1 - \exp(-|\mathcal{N}|(u + v))}{|\mathcal{N}|(u + v)}
\tag{44}
$$

where $|\mathcal{N}|$ is the number of points in the interaction neighborhood.

Comparing Eqns. (43) and (42) reveals two differences : there is an extra 1 in the death term but more importantly, the form of the game term has changed. To examine the second difference, we note that $h(u, v)$ converges to 1 as $(u+v) \to 0$ and decreases to 0 as $(u+v) \uparrow \infty$. When h is close to 0 (which occurs for example if $|\mathcal{N}|$ is large) the game term is almost what it was before. When h is close to 1, the game term is almost a in the first equation and d in the second equation, reflecting the fact that when the density is small most individuals are isolated and think that the universe consists exclusively of their own type.

The changes in the limiting equation bring about a drastic change in the behavior of the reaction-diffusion equation and the associated dynamical system and that is the dynamical system *now has a globally attracting fixed point*. There are other systems that can be obtained based on slightly different assumptions about the limiting behaviors but they seem to share the existence of an attracting fixed point.

159

6 From level to level

Physics contains many different models, which might describe different aspects of the very same physical system. A gas is composed of molecules, but it also obeys the laws of fluid mechanics. Or again, a fluid near its critical point is a bunch of molecules, but they can also be described by a scale-invariant field theory. The microscopic rules are time-reversible whereas the macroscopic world has irreversibility. The common-sense physics that we as humans need to know in dealing with our generally low-viscosity world is quite different from the physics a bacterium would need to know to deal with its small-scale high-viscosity world, though both are described by the same laws at a more fundamental level. In each of these cases, there is a tension between two levels of description. For each situation, different laws, formulations, conceptualizations, theories and experiments apply at each of the two levels. Clearly, it is important to know how the different realities caught by the different models fit together. Statistical mechanics provides methods to relate the macroscopic behavior of a system with a large number of degrees of freedom, to the laws governing its microscopic behavior.

As is clear from the examples discussed in this lecture, similar questions are being asked in basic ecology and in the context of applied problems like global climate change. It is hoped that what has worked in physics might also work here. In order to facilitate this discussion, we construct a toy model based on the work of Durrett & Levin (1994) which captures several essential features of spatial population dynamics. We then draw parallels between efforts to study similar systems in condensed matter physics and fluid dynamics to motivate questions to be asked about our model and possible ways of answering them.

6.1 Lagrangian population dynamics

In the interacting particle system discussed in §5, population is discrete, time is continuous and space is discrete. We extend this model by making space continuous as well i.e., by removing the restriction that animals can only live on the lattice and allowing them to move as points in space. The dynamics are formulated as follows :

Migration. Each individual performs brownian motion. We specify the diffusion coefficient characterizing p(x,t), the PDF of the particle.
Deaths due to crowding. Each individual at x at time t dies at rate $\kappa(\eta_t(x) + \zeta_t(x))$.
Game step. Each hawk experiences a birth (or death) rate of $ap_t(x) + b(1 - p_t(x))$ while each dove experiences a birth (or death) rate of $cp_t(x) + d(1 - p_t(x))$.

Here, $\eta_t(x), \zeta_t(x)$ is the number of hawks and doves respectively in a circular region of radius R (the interaction radius) centered at x. As before, $p_t(x) = \eta_t(x)/(\eta_t(x) + \zeta_t(x))$.

This model is appealing on several counts :

160

- *Higher spatial resolution.* By going to a grid-free treatment of space, the resolution of features is limited by the number of particles and not the discreteness of space.

- *Advection.* It is straightforward to advect the animals by fluid flow to mimic situations in marine environments. More generally, we can view the flow as the tendency to perform taxis following the gradient of some environmental variable or to model seasonal migrations.

- *Social behavior.* It is quite easy to incorporate social behavior by causing the animals to move towards/away from each other as they try to optimize the number of neighbors or change tactics depending on the density of predator/prey.

6.2   Coarse graining

We want to understand how the macroscopic behavior of our model changes as the microscopic parameters are changed. By macroscopic behavior, we mean (a) the equilibrium behavior for the average densities of hawks and doves and (b) the dynamics of relaxation to equilibrium. Our system is dominated by fluctuations and so the study of our system is essentially a study of fluctuations. Monte Carlo simulations have been used successfully to study fluctuating systems in physics (Mouritsen, 1984 and Binder & Heermann, 1992). Below, we briefly look at how this method is used for the Ising model. Our system is conceptually not different from the Ising model and so Monte Carlo methods can be applied to study the equilibrium and the approach to equilibrium in our system as well. In the vicinity of phase transitions in the Ising model, correlation lengths and relaxation times diverge. This challenges the Monte Carlo method in the vicinity of phase transitions and one must resort to other methods such as the renormalization group. Our system has phase transitions as well with the associated large correlation lengths and long relaxation times and in order to study these behaviors, other techniques have to be looked into.

6.2.1   Markov chain Monte Carlo method for the Ising model

The 2-D Ising model consists of a square lattice and associated with each point of the lattice is a magnetic spin, a number s that is either 1 or -1. The system's Hamiltonian in the absence of an externally imposed magnetic field is

$$H = -J \sum_{<i,j>} s_i s_j, \quad J > 0$$

where J is the exchange energy and the sum is taken over all nearest-neighbor pairs. Let Ω represent a microstate i.e., a particular assignment of ± 1 to lattice sites. The probability of a certain microstate is

$$p(\Omega) = \frac{1}{Z} \exp[-\beta H(\Omega)].$$

161

Here, $\beta = 1/k_B T$ where k_B is Boltzmann's constant, and T is the temperature and Z is the partition function given by

$$Z = \sum_{\{\Omega\}} \exp[-\beta H(\Omega)]$$

where the sum is over all the microstates. Given the probability distribution of the microstates, the thermodynamic value of a measurable physical quantity, $f(\Omega)$, is obtained in the canonical ensemble as

$$< f > = \sum_{\{\Omega\}} f(\Omega) p(\Omega). \tag{45}$$

Equation (45) constitutes the formal connection between the microscopic and macroscopic worlds.

The idea behind Monte Carlo methods is to approximate the above sum over all configurations by a sum over a set of representative configurations. The simplest possible procedure would involve an unbiased choice of configurations (simple sampling). In this case, we have

$$< f > = \frac{\sum_{i=1}^{M} \exp[-\beta H(\Omega_i)] f(\Omega_i)}{\sum_{i=1}^{M} \exp[-\beta H(\Omega_i)]}.$$

This procedure is inefficient since the Boltzmann weights ($\exp[-\beta H]$) vary over orders of magnitude and we are wasting the effort used to compute the configurations that do not contribute much to the partition function. The thing to do is to bias the choice of configurations so that we only pick out configurations that make a big contribution to the partition function. The simple and most natural choice is to pick a configuration Ω with probability $\exp[-\beta H(\Omega)]$. Then,

$$< f > = \frac{1}{M} \sum_{i=1}^{M} f(\Omega_i).$$

The problem is, of course, to find a procedure which practically realizes this so-called importance sampling. Metropolis advanced the idea not to choose the successive states $\{\Omega\}$ independently of each other, but to construct a Markov process where each state is constructed from a previous state via a suitable transition probability. One such way is the famous Metropolis algorithm. Using it, we can determine the equilibrium values of quantities such as magnetization. Further, we can think of the evolution of the Markov chain as time evolution of the system. This dynamic interpretation of the importance sampling Monte Carlo method allows us to study the relaxation to equilibrium. Typically, one uses $10^2 - 10^4$ Markov chains each of length $10^3 - 10^5$ to get reasonable statistics. This method is very effective at higher temperatures where (a) the correlation lengths are small and the size of the system can be made much larger than the correlation length and (b) the relaxation times are short in comparion with the length of the Markov chains and so equilibrium is not hard to reach.

In our model, we have a dynamic process which can be interpreted as a Markov chain and so it is straightforward to analyze our system using the Monte Carlo method.

162

6.2.2 Finite size effects

In scenario 3 in §5, the mean-field approximation predicts global extinction whereas the real system has non-zero average hawk and dove densities. The discrepancy arises because the mean-field approximation assumes that all hawks and doves are within the interaction radius of each other. If the interaction radius, $R \sim L$ where L is the system size, the mean-field approximation would hold. The question of interest then is how the macroscopic behavior changes as we increase L and approach the limit $R \ll L$. This question can be answered with the Markov chain Monte Carlo method. One fixes R and considers successively larger values of L and tracks how the macroscopic behavior changes. By looking at the scaling, behavior for $R \ll L$ can be predicted.

6.2.3 Broken ergodicity, critical slowing down and phase transitions

The realization of the Markov chain Monte Carlo method will provide thermodynamic equilibrium provided that the process is ergodic. Ergodicity requires that any one configuration of the Markov chain is accessible from any other configuration via a finite number of transitions. There are two ways in which the ergodicity hypothesis can be violated. First, the phase space might be reducible into several ergodicity classes and second, the relaxation times τ_r might be very much longer than the length of the Markov chain τ_{obs} in which case ergodicity is effectively broken. We examine both of these cases.

A. Reducible phase space

Consider scenario 3 in §5. If there is global extinction, then there is no way for the system to populate itself. Hence, the global extinction part of the phase space is accessible from the rest of the phase space but the reverse is not true. If we consider a system large enough, then the likelihood of global extinction decreases and we can be reasonably certain that the system is ergodic.

Alternatively, consider the behavior of our model for

$$
\begin{array}{ll}
a = 0.7 & b = 0.4 \\
c = 0.4 & d = 0.7
\end{array}
\tag{46}
$$

This represents two populations interacting where each helps its own kind more than the other. Owing to the symmetry in the problem, there are two possible stable end-states viz., either the hawks survive or the doves survive. Which state is picked out depends on the initial condition. If we start with a large number of hawks and few doves it will be very hard to reach the part of phase space where doves dominate. Hence, we effectively have two reducible ergodicity classes and we can study each one in turn using the Monte Carlo method.

163

B. Extremely long relaxation times

Again, consider the model (46). Now, we are interested in the time it takes for the system starting from a particular initial condition to reach either of the equilibria. If the initial state had more hawks than doves, then the hawks will quickly dominate and lead to extinction of doves. Clearly, if we started with even more hawks initially, we would reach the extinction of doves even faster. Consider, then what happens if we start with roughly equal number of hawks and doves in a really large domain. Owing to random fluctuations in the system, locally either the hawk or dove population will dominate and form clusters. However, once this state is reached, evolution becomes exceedingly slow as the dynamics of cluster-cluster interaction is very slow. On a much slower time-scale then, the clusters will grow which will be reflected in increasingly large correlation lengths. This is called "critical slowing down" and is observed in biopolymers, anisotropic magnets and binary metallic alloys. This phenomenon is associated with the kinetics of phase transitions and is covered in the review volumes by Binder (1986, 1984). There is reason to believe that the time to reach equilibrium behaves as $t \sim |x|^{-\alpha}$ where x is some appropriate distance to the separatrix dividing the basins of attraction of the two equilibria. Of particular interest would be the calculation of the 'dynamic critical exponent' α.

6.3 Stochastic differential equations

Below, we look at how stochastic differential equations have been used to study the turbulence problem. This suggests that a similar approach might be useful in studying our model as well. First, in order to set up the notation, consider the stochastic differential equation

$$dX_t = a(t, X_t)dt + b(t, X_t)\xi_t dt$$

where X_t is the random variable whose change is given by a deterministic or averaged drift term $a(t, X_t)$ perturbed by a noisy diffusive term $b(t, X_t)\xi_t$, where the ξ_t are standard Gaussian random variables for each t and $b(t, X_t)$ is a state-time dependent intensity factor. This symbolic differential is interpreted as an equation

$$X_t(\omega) = X_{t_0}(\omega) + \int_{t_0}^{t} a(s, X_s(\omega))ds + \int_{t_0}^{t} b(s, X_s(\omega))dW_s(\omega) \qquad (47)$$

for each sample path ω where W_t is the corresponding Wiener process.

6.3.1 Turbulence problem

In turbulence, one studies homogenous isotropic turbulence in a 3-D periodic box as a model for studying features of real turbulence. If we are studying turbulence in the wake of a car, for instance, this model of turbulence only applies on scales much smaller than the dimension

of the car where the dynamics and mechanisms of how the turbulence was generated do no longer matter. The governing equations are the incompressible Navier-Stokes equations :

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\frac{1}{\rho}\nabla p + \nu \nabla^2 \mathbf{u} + \mathbf{F}$$

$$\nabla \cdot \mathbf{u} = 0$$

where $\mathbf{u}(\mathbf{x}, t)$ is the fluid velocity, ρ is the density, p is the pressure, ν is the molecular viscosity and $\mathbf{F}$ is the force. The force models the effect of the outside world on the flow in the box. In absence of this force, the turbulence will transfer energy to ever smaller scales until it is dissipated by molecular viscosity and the fluid would eventually come to rest. Usually, one considers Gaussian, white-in-time random force in a narrow wavenumber band.

6.3.2   Local extinction scenario

Consider the behavior of our model for scenario 3 discussed in §5.

$$a = -0.6 \quad b = 0.9$$
$$c = -0.9 \quad d = 0.7$$

The behavior of the lagrangian model is qualitatively not different from the interacting particle system model. In absence of migration of particles, the populations collapse. Migration provides the mechanism for recolonization and the cycle can then continue.

Let us consider a system of size L and interaction radius R such that $R \sim L$ and L is large. In this case, the behavior of the system is well approximated by the mean-field model :

$$
\begin{aligned}
\frac{du}{dt} &= u\left\{a\frac{u}{u+v} + b\frac{v}{u+v} - k(u+v)\right\} \\
\frac{dv}{dt} &= v\left\{c\frac{u}{u+v} + d\frac{v}{u+v} - k(u+v)\right\}
\end{aligned}
\tag{48}
$$

Now, consider that we have an assembly of such systems and we want to model the effect of the other such systems on one of them, not knowing the states of the other systems. The only coupling between these systems is through migration i.e., net loss or gain of particles as they move between systems. This can be modeled through a white-noise term in the above ODE leading to

$$
\begin{aligned}
dU_t &= U_t\left\{a\frac{U_t}{U_t+V_t} + b\frac{V_t}{U_t+V_t} - \kappa(U_t+V_t)\right\}dt + \sigma^1(U_t)dW_t^1 \\
dV_t &= V_t\left\{c\frac{U_t}{U_t+V_t} + d\frac{V_t}{U_t+V_t} - \kappa(U_t+V_t)\right\}dt + \sigma^2(V_t)dW_t^2
\end{aligned}
\tag{49}
$$

Care has to be taken to make sure that the densities do not become negtive.

6.3.3 Ito or Stratonovich ?

The difference between Ito and Stratonovich calculi comes from interpreting the latter integral in (47) when approximating the integral with a discrete sum. In Ito calculus, the integrand $b(s, X_s(\omega))$ is evaluated at the left hand of the discretization intervals whereas in Stratonovich calculus, the evaluation point is chosen to be the middle of the discretization interval. Ito calculus has the martingale property which turns out to be crucial in stochastic analysis quite often. Stratonovich calculus does not have the martingale property but it does enjoy similar properties to those of classical calculus which the Ito calculus does not have. They are both mathematically valid formulations of stochastic differential equations and we need to decide which interpretation is appropriate for our problem.

Stratonovich calculus is the appropriate one when the white noise is used as an idealization of a smooth real noise. In our problem, this is not the case. Moreover, we have discrete population sizes and the dynamics are nonanticipative and hence, we say that the *Ito interpretation is the appropriate one for our problem.*

6.3.4 Numerical solution

For direct simulation of trajectories, strong schemes are required. Kloeden & Platen (1992) and Kloeden, Platen & Schurz (1994) give an in-depth introduction to numerical schemes for the integration of stochastic differential equations. Of crucial importance is the steady probability density function $p^*(u, v)$. There is a Fokker-Planck equation for the evolution of $p(u, v, t)$ and there is reason to believe that for all initial conditions, $p(u, v, t)$ tends to the same steady-state $p^*(u, v)$ after some time. Weak schemes can be used to calculate this distribution function.

6.3.5 What does SDE say about our model ?

Once we have an efficient method for computing $p(u, v)$, then we can practically change the microscopic parameters and observe its effect on the stationary probability distribution. Hence, even though adding white noise to the mean-field model may not accurately reproduce the equilibrium states or the dynamics of relaxation to equilibrium, there is a good chance that it will mimic the relation of microscopic laws to macroscopic behavior.

References

[1] BINDER, K., ED. 1986 *Monte Carlo methods in statistical physics* Topics in Current Physics, vol. 7, Springer-Verlag.

[2] BINDER, K., ED. 1984 *Applications of the Monte Carlo method in statistical physics* Topics in Current Physics, vol. 36, Springer-Verlag.

[3] BINDER, K. & HEERMANN, D. W. 1992 *Monte Carlo simulation in statistical physics : An introduction* Springer Series in Solid-State Sciences, vol. 80, Srpinger-Verlag.

[4] DURRETT, R. & LEVIN, S. 1994 The importance of being discrete (and spatial). *Theor. Popul. Biol.* **46**, 363-394.

[5] GUERON, S. & LEVIN, S. 1993 Self-organization of front patterns in large wildebeest herds. *J. Theor. Biol.* **165**, 541-552.

[6] GUERON, S. & LEVIN, S. 1995 The dynamics of group formation. *Mathematical Biosciences* **128**, 243-264.

[7] IWASA, Y. & LEVIN, S. 1995 The timing of life history events. *J. Theor. Biol.* **172** 33-42.

[8] KADANOFF, L. P. 1993 *From order to chaos* World Scientific.

[9] KLOEDEN, P. E. & PLATEN, E. 1992 *Numerical solution of stochastic differential equations* Applications of Mathematics, vol. 23, Springer-Verlag.

[10] KLOEDEN, P. E., PLATEN, E. & SCHURZ, H. 1994 *Numerical solution of SDE through computer experiments* Universitext, Springer-Verlag.

[11] LEVIN, S. 1992 The problem of pattern and scale in ecology. *Ecology* **73**(6), 1943-1967.

[12] MOURITSEN, O. G. 1984 *Computer studies of phase transitions and critical phenomena* Spinger Series in Computational Physics, Springer-Verlag.

[13] MURRAY, J. D. 1993 *Mathematical Biology* Biomathematics, vol. 1, Springer-Verlag.

[14] OKUBO, A. 1980 *Diffusion and Ecological Problems : Mathematical Models* Biomathematics, vol. 10, Springer-Verlag.

[15] SINCLAIR, A. R. E. 1977 *The American Buffalo, A Study of Resource Limitation of Population* The University of Chicago Press, Chicago, IL, plate 3.

Notes compiled by Amar Gandhi.

Ekman drift modulation as a factor in connections between coastal eddy systems and fisheries recruitment in the Florida Keys.

Claes Rooth,
School of Marine and Atmospheric Science, University of Miami.

Persistent oceanic eddy/recirculation patterns have long been recognized for their possible role in isolated island ecosystems as temporary storage reservoirs for pelagic larvae and juveniles which might otherwise be lost from the local system through large scale advection processes. While such effects are affected by local geographic and climatological circumstances, general characteristics of the retention process can be identified, such as conditions under which accretion of biota from the outside may occur into the retention domain , and whether preferred exit paths from retention "reservoirs" can be identified . Such questions form the focus of a subproject within the Southeast Florida and Caribbean Recruitment study (SEFCAR),. which seeks to explore the impact of a distinct regional ocean circulation climatology in the Straits of Florida and Caribbean regions on the recruitment processes of a variety of fish and crustacean species.

The general physical influences on the Florida Keys recruitment processes have been described by Lee & al. (1992) as follows: The Gulf Stream/Florida Current provides an ever-present strong advection field off-shore of the Florida Keys, but its precise path oscillates widely and irregularly in response to the quasi-periodic Gulf Stream intrusion (known as the Loop Current) into the Gulf of Mexico. The normal flow pattern in the entrance to the Straits of Florida takes the form of a wide cyclonic loop, the Tortugas Gyre, with a counterflow along the westernmost section of the Florida Keys. A smaller, and somewhat less persistent loop circulation tends to form along the middle Keys in a region of slightly wider shelf topography, the Pourtales terrace. Occasional cutoff of the Loop Current at its neck between the Yucatan Peninsula and the Florida Keys leads to a temporary disappearance of the recirculation cells, and advection of the core waters of the Pourtales gyre and at least part of the Tortugas Gyre through the Straits of Florida, these gyres provide a significant albeit intermittently interrupted regional retention probability. In combination with the climatological wind forcing, they provide a mechanism for enhancing return probabilities to the inshore reef tracts through the interaction between the wind induced (Ekman) drift effects and the eddy circulations.

Three aspects of these interactions will be discussed here, viz. (1) the modification of the Ekman drift by properties of the surface current field, (2) the resulting convergence patterns within eddy retention zones, and (3) the onshore transport effect associated with cyclonic coastal eddies under coastal downwelling conditions.

Oceanic wind systems are in general not conducive to direct generation of baroclinic circulation responses in the upper ocean layers, of a character which elicits strong biological response effects. Their horizontal scales are too far removed from the so called deformation radius associated with near surface trapped circulations (R_d=c/f, where c is the long gravity wave speed associated with near surface motion modes, and f is the Coriolis parameter. Values of the order of 5-10 km are typical in mid and high latitude situations). The primary biological impacts of direct wind forcing are therefore associated either with edge effects, as in coastal upwelling, or with modifications of the general upper layer stratification and nutrient supplies by vertical transport processes. It turns out, however, that surface responses on frontal scales are induced by the structure of surface currents even when the wind forcing has much larger horizontal scales.

Niiler (1969) observed that mesoscale Ekman drift convergence will be induced by the vorticity fields in surface currents, in proportion to their strength relative to the

Coriolis parameter, i.e. to the Rossby number of a current jet. He did not recognize, though, that an equally important effect arises because the surface current modifies the shear in the near surface wind. Bye (1986) introduced the term "understress" for the latter effect, and suggested that it might introduce a significant damping rate for oceanic surface currents. Dewar and Flierl (1987) explored the impact of such stress feedback effects on isolated current rings, with emphasis on their dissipative effects. Upwelling effects implicitly related to the latter were explored by Rooth and Xie (1992, RX hereafter)) and by Rooth et al. (1994). RX also showed that the understress effect and the Ekman drift convergence due to surface vorticity gradients can be unified in terms of equivalent stress curl effects because the latter can also be interpreted as induced by boundary layer vorticity transport by the Ekman drift.
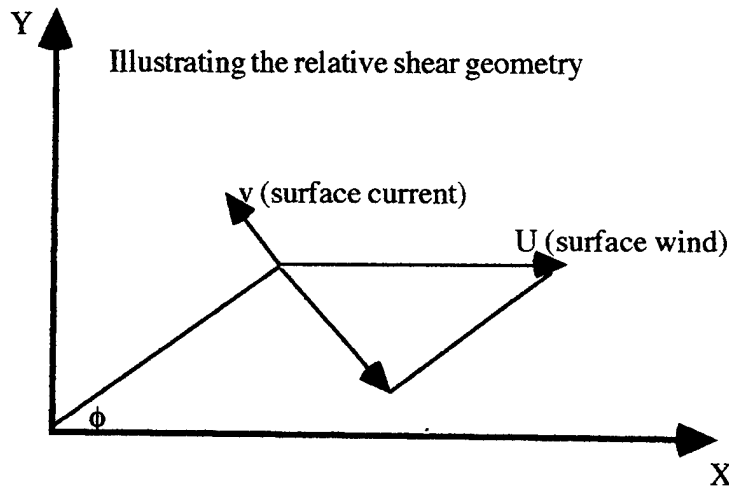
In the above illustrated relative geometry, the effective wind shear in the current direction is $U\sin\phi\text{-}v$, and the total shear velocity $U+v\sin\phi$ (to order v/U). The total kinematic stress in the current direction (determining the cross current Ekman drift) is to order v^2 seen to be:

$$\tau_c = - c_d(U^2\sin\phi - Uv(1+ \sin^2\phi)) \qquad (1)$$

where c_d is the surface drag coefficient normalized by the density ratio between the media (we shall here assume that $c_d\approx2\times10^{-6}$). For a straight oceanic jet, the surface vorticity is given by the current shear, v_n, and the effective stress curl is found to be:

$$\text{curl } \tau_c = -c_dU(v_n(1+ \sin^2\phi) + f^{-1}Uv_{nn}) \qquad (2)$$

where the second term represents the Ekman drift convergence effect identified by Niiler (1969). Note that although this term is likely to be instantaneously dominant under many circumstances, it vanishes for a time variable wind with zero mean, while the first term is only weakly sensitive to wind direction and thus cumulative in its effects.

In considering biological impacts of near surface dynamic processes in the oceans, Lagrangian temporal persistence of the phenomena becomes a major concern, since full evolution of biological impacts may require times typically counted in days or more. That is, homogeneity under advection in a frontal jet is critical in simple models of slow

responses in a fast flowing current. As an example, water in the Gulf Stream core with a current speed of 1-2 m/sec is displaced of the order of 100 km in a day.

The case of current rings is very different because of the recirculation and resultant trapping of waters that occurs in these features. In an isolated current ring, where the tangential velocity vanishes at the outer boundary, the area integrated vorticity must vanish. For a symmetric ring, a streamline must then exist where the vorticity vanishes, separating a core region with vorticity of the rotation sign of the ring, and an outer one with the opposite vorticity sign. The under-stress effect will induce convergence in the cyclonic vorticity zone, and divergence in the anticyclonic one, and thus induce concentration or dispersal tendencies in different radial domains of a ring.

The radial Ekman transport, M_r, is:

$$M_r = - (f+v_n)^{-1} c_d(U^2\sin\phi - Uv(1+ \sin^2\phi)) \tag{3}$$

The effect of the mean ambient Ekman drift (the first term in III) is to distort the surface circulation, setting up a retention zone which is displaced from the eddy center as illustrated below in a surface stream function section, taken parallel to the wind and through the center of an axisymmetric cyclonic ring. Its average around a streamline vanishes. In the figure, the slanted lines are a stream function representation of the ambient Ekman transport. Neglecting for the present the correction due to the ring vorticity, the Ekman drift modified stream lines are the horizontal projections of the intersections between a set of slanted planes and an axisymmetric bowl defined by its cross section in the figure, and the closed resultant stream lines are generated by the set of intersections between the illustrated ones.
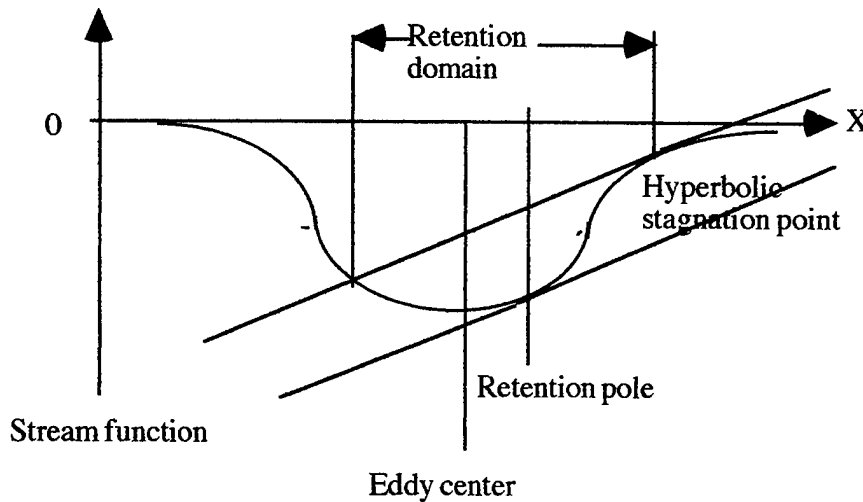
Assuming that the retention zone stream lines are still approximately circular, and noting that the radial Ekman drift correction due to the ring vorticity vanishes at the core boundary, we can estimate the total Ekman drift convergence rate into the core region as:
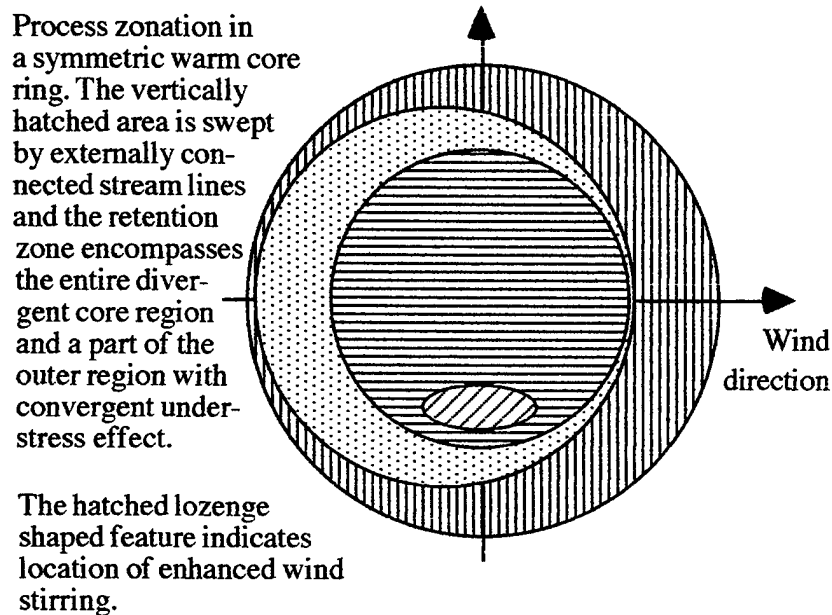
$$\oint M_r r d\phi = 3\pi r c_d f^{-1} Uv \tag{4}$$

and the corresponding mean vertical velocity inside the ring core region as:

170

$$\overline{w} = \frac{3c_dUv}{fr} \approx 3c_dUR_o \tag{5}$$

Rossby numbers (R_o) for well developed rings are typically of order 0.1 to 0.5. We choose 0.2 (appropriate for the Florida Keys eddies) for an example computation, take the density normalized drag coefficient as $2x10^{-6}$, and adopt a moderate wind speed of 7,5 m/sec (15 knots), to yield an up(down)welling rate of $8x10^{-6}$m/sec, or approximately 0.7 m/day in a warm (cold) core ring. With ring life times of the order of a month, this provides significant mass displacements between the radial ring zones.

Lateral ventilation of the surface layer along the trajectories outside the retention zone does occur with the mean Ekman drift. For a symmetric ring under steady wind conditions, the entire Ekman flux that hits the eddy boundary is shunted around the retention zone through a crescent shaped region which encompasses approximately one half of the total eddy area, as illustrated below:

Process zonation in a symmetric warm core ring. The vertically hatched area is swept by externally connected stream lines and the retention zone encompasses the entire divergent core region and a part of the outer region with convergent understress effect.

The hatched lozenge shaped feature indicates location of enhanced wind stirring.

Wind direction

If the wind direction should reverse, something like half of the affected surface area will be included in the restructured retention zone. Thus slowly variable winds may in fact engender an effective eddy diffusivity between the eddy surface layer and its surroundings.

In addition to the understress induced convergence effects, there is also a modulation of the wind induced mixing processes. Conventional approximations suggest proportionality to the cube of the friction velocity, and thus a substantial enhancement of the wind stirring in the region of maximal opposing current speed, the center of which is indicated in the figure by the diagonally hatched lozenge shape.

Returning our attention to the Florida Keys situation, we have there examples both of the current jet (Gulf Stream) and ring current interaction situations (Tortugas and Pourtales gyres). Both of the gyre objects are cyclonic in nature, and thus characterized

by convergence effects in their core regions. As cyclonic recirculation zones along a climatological downwelling boundary, they provide the possibility of significant storage effects for maturing larvae and early juvenile stages of a number of locally important fish and crustacean species. A rather unique aspect of the Florida Straits climatology is however the juxtaposition of a strong boundary current and opposing winds over much of the year. As discussed by Rooth et al. (1994), this leads not only to the prospect of local retention and return to the inshore region of progeny of reef populations, but also to a prospect for capture by the coastal eddies of various flotsam, whether biological or inanimate, which is carried by the Gulf Stream jet. Once captured, such objects can be carried onshore by the quasi-persistent coastal gyres, to be preferentially delivered onshore in their downstream quadrants (relative to the Gulf Stream flow). A surface layer strip of several kilometers with is estimated to be diverted from the stream into the Pourtales Gyre circulation along the roughly 40 km downstream width of the latter.

Because of the somewhat ephemeral character of these inshore gyres, particularly of the smaller Pourtales Gyre (off the middle Keys), a mechanism is here provided for substantial variability in local recruitment success, dependent on the relative timing of spawning events and the vagaries of the gyre structures. Further research is needed particularly on how the effective thickness of the Ekman drift layer and the vertical distribution habits of various organisms interact to modify the transport rates. One may in particular expect that diel migrations of zooplankton. over a depth range which exceeds the Ekman drift profile scale, should introduce a strong relative dispersal factor in trophic level interactions whenever the "mixed layer is poorly mixed". New techniques for surface drift current mapping based on VHF radio backscatter from surface waves, combined with ins situ continuous sampling techniques along ship transects are being applied in the SEFCAR project, and will hopefully in time lead to clarification of some of these problems.

References.:

Bye, J. A. T, 1986:Momentum exchange at the sea surface by wind stress and understress. Q. J. Roy Met. Soc. 112, 501-510

Dewar, W. K. and G. R. Flierl, 1987: Some Effects of the Wind on Rings. J. Phys. Oceanography, 17, 1653-1667.

Lee, T. N., C. Rooth, E. Williams, M. McGowan, A Szmant and M. E. Clarke, 1992: Influence of Florida Current, gyres and wind driven circulation on transport of larvae and recruitment in the Florida Keys coral reefs. Continental Shelf Research, 12, 971-1002

Niiler, P. P., 1969: On the Ekman Divergence in an Oceanic Jet. J. Geophys. Res. 74, 7048-7052

Rooth, C. and L. Xie, 1992: Air-sea boundary layer dynamics in the presence of mesoscale surface currents. J. Geophys. Res.,. 97, pp 14,431-14,438.

Rooth, C. G., L. Xie and T. N. Lee, 1994: Effect of wind direction on the surface transport and retention patterns in a coastal eddy. In press: J. Physical Oceanography, 24.

Survival: What is it and when does it occur?

Chris Cosner

Dept. of Mathematics

Univ. of Miami

Coral Gables, FL 33124

One of the basic problems in ecology is that of determining whether a population, species, or community can be expected to survive. A common theoretical approach is to build and study mathematical models for the population dynamics of the species of interest and their interactions with the environment and each other. If the population dynamics are simple and well-understood, and the physical environment is homogeneous, then there may be a globally attracting equilibrium for the model with all equilibrium populations or densities at high levels. In general, however, there are some major problems with this approach. First, even simple systems may have complicated dynamics such as limit cycles or strange attractors, *i. e.*, chaos. Second, the influence of the physical environment may be complex and/ or erratic, so that a complete description of the population dynamics is not technically feasible. Third, the interactions of a population with the environment or other populations may be known only approximately. Thus it is worthwhile to consider alternatives to the existence of a globally attracting equilibrium. Three of those are compressivity, permanence/ uniform persistence, and practical persistence.

A model which can be viewed as a dynamical system is said to be *uniformly persistent* if it has a globally attracting set which is bounded below in every component by a positive number or density. It is said to be *permanent* if the attractor is also bounded above. Usually a system must be autonomous or periodic in time for the idea of permanence to apply. Typically permanence is established for a model of m species by showing that any subsystem with $m - 1$ species at steady state is unstable with respect to the remaining species; *i.e.*, the remaining species can invade if introduced at low densities. A review of the theory is given in Hutson and Schmitt (1992); see also Cantrell, Cosner, and Hutson (1991, 1993).

A model which preserves an ordering of the initial data is said to be *compressive* if there exists a minimal steady state which is stable from below and a maximal steady state which is stable from above. The crucial requirement is that the system be order-preserving. That will typically be the case for continuous time models of a single species, two competitors, or an arbitrary number of mutualists, but will fail for predator-prey systems or for three or more competitors. The idea of compressivity is discussed by Hess (1991) in the context of reaction-diffusion models. Criteria for systems of ordinary differential equations to be order-preserving are given by Smith (1988); often those criteria will extend to reaction-diffusion systems.

In principle, *practical persistence* simply means uniform persistence or permanence with some explicit lower bounds for the globally attracting positive set. In practice, practical persistence is typically established by a collection of *ad hoc* techniques which include the use of multiple Lyapunov functions and comparisons with simpler subsystems or single species models. There is little general theory (yet) for practical persistence, but related ideas are discussed in Cao and Gard (1993) and have been used in the analysis of reaction-diffusion systems in Cosner and Lazer (1984) and Cantrell, Cosner, and Hutson (1993). A

detailed treatment is given in Cantrell and Cosner (preprint). A major advantage of the *ad hoc* methods used to establish practical persistence is that they may be used in situations where the coefficients in the model are only approximately known or have complicated (but bounded) time dependence.

To see how these ideas apply or fail, consider the reaction–diffusion models

$$\frac{\partial u}{\partial t} = D\nabla^2 u + (a - bu)u \tag{1}$$

$$\frac{\partial u}{\partial t} = D_1\nabla^2 u + (a - bu - cv)u$$
$$\frac{\partial v}{\partial t} = D_2\nabla^2 v + (d - eu - fv)v \tag{2}$$

$$\frac{\partial u}{\partial t} = D_1\nabla^2 u + (a - bu - cv)u$$
$$\frac{\partial v}{\partial t} = D_2\nabla^2 v + (-d + eu)v \tag{3}$$

$$\frac{\partial u}{\partial t} = D_1\nabla^2 u + [a(t,u,v) - b(t,u,v)u - c(t,u,v)v]u$$
$$\frac{\partial v}{\partial t} = D_2\nabla^2 v + [d(t,u,v) + e(t,u,v)u + f(t,u,v)v]v \tag{4}$$

where the spatial domain is a bounded region Ω and $u = v = 0$ on $\partial\Omega$. Model (2) is for competition; (3) and (4) are predator–prey models.

For (1) there will be a unique positive equilibrium density $u = \theta(D, a, b)$ which is globally attracting among positive solutions whenever $\sigma_0 > 0$, where σ_0 is the largest eigenvalue of

$$D\nabla^2\psi + a\psi = \sigma\psi \quad \text{in } \Omega$$
$$\psi = 0 \quad \text{on } \partial\Omega.$$

Furthermore, if we normalize ψ by $\max\psi = 1$, then $(\sigma_0/b)\psi \leq \theta(D, a, b) \leq a/b$ (see for example Cosner and Lazer, 1984).

The existence of a globally stable positive equilibrium in (2), (3), or (4) has been established only under quite restrictive hypotheses about the coefficients. On the other hand, the ideas of permanence, compressivity, and practical persistence all apply to (2) provided the largest eigenvalues of

$$D_1\nabla^2\psi_1 + [a - b\theta(D_2, d, f)]\psi_1 = \sigma\psi_1$$
$$D_2\nabla^2\psi_2 + [d - e\theta(D_1, a, b)]\psi_2 = \sigma\psi_2 \quad \text{in } \Omega$$

$$\psi_1 = \psi_2 = 0 \quad \text{on } \partial\Omega .$$

are both positive; see Cantrell, Cosner, and Hutson (1991), Hess (1991) and Cosner and Lazer (1984) respectively. For (3) and (4) compressivity fails, since the systems are not

order–preserving. It seems difficult or impossible to establish practical persistence in (3) because of the difficulty in obtaining upper bounds on v. Permanence will hold for (3) if the largest eigenvalues of

$$D_1 \nabla^2 \psi_1 + a\psi_1 = \sigma\psi_1$$
$$D_2 \nabla^2 \psi_2 + [-d + e\theta(D_1, a, b)]\psi_2 = \sigma\psi_2 \quad \text{in } \Omega$$

$$\psi_1 = \psi_2 = 0 \quad \text{on } \partial\Omega.$$

are both positive; see Cantrell, Cosner, and Hutson (1991). On the other hand, permanence in (4) can be established only if the time dependence in the coefficients is periodic or at worst almost periodic and the coefficients are explicitly known. However, even if the only information available on the coefficients is a collection of upper and lower bounds, it is sometimes possible to establish practical persistence. The key point is that a single reaction–diffusion equation is always order–preserving, and even a reaction–diffusion inequality will retain some order–preserving properties. Suppose that in (4) we have $\underline{a} \leq a(u,v,t) \leq \overline{a}$, $\underline{b} \leq b(u,v,t) \leq \overline{b}$, etc. Then for any solution (u, v), u must satisfy

$$\frac{\partial u}{\partial t} \leq D_1 \nabla^2 u + (\overline{a} - \underline{b}u)u \quad,$$

so, by comparison with (1), we have $u \leq \overline{a}/\underline{b}$ for t large. When enough time has passed to realize that bound on u, we have

$$\frac{\partial v}{\partial t} \leq D_2 \nabla^2 v + (\overline{d} + \overline{e}(\overline{a}/\underline{b}) - \underline{c}v)v$$

so that comparison with (1) shows that eventually $v \leq M \equiv [\overline{d} + \overline{e}(\overline{a}/\underline{b})]/\underline{c}$. [This is the point where the argument for practical persistence breaks down for (3).] Having obtained an asymptotic upper bound on v, we can obtain an asymptotic lower bound on u provided that $\underline{a}$ is large enough and $\overline{c}$ or M is small enough. For t large,

$$\frac{\partial u}{\partial t} \geq D_1 \nabla^2 U + (\underline{a} - \overline{c}M - \overline{b}u)u$$

so that if the largest eigenvalue of

$$D_1 \nabla^2 \psi + (\underline{a} - \overline{c}M)\psi = \sigma\psi \quad \text{in } \Omega$$

$$\psi = 0 \quad \text{on } \partial\Omega$$

is positive, we have $u \geq \theta(D_1, \underline{a} - \overline{c}M, \overline{b})$ for large t. As previously noted, this lower bound can be estimated in terms of ψ and the eigenvalue. Finally, we can use the asymptotic lower bound on u to generate an asymptotic lower bound on v. For t large,

$$\frac{\partial v}{\partial t} \geq D_2 \nabla^2 v + (\underline{d} + \underline{e}\theta(D_1, \underline{a} - \overline{c}M, \overline{b}) - \overline{f}v)v \quad,$$

175

so $v \geq w$ where w has the same initial data as v at some large value of t and

$$\frac{\partial w}{\partial t} = D_2 \nabla^2 w + (\underline{d} + \underline{e}\theta(D_1, \underline{a} - \bar{c}M, \bar{b}) - \bar{f}w)w \tag{5}$$

with $w = 0$ on $\partial\Omega$. Although (5) has a spatially varying linear growth term, it behaves in much the same way as (1). Specifically, equation (5) will have a unique globally stable positive equilibrium ρ provided that the largest eigenvalue of

$$D_2 \nabla^2 \psi + (\underline{d} + \underline{e}\theta(D_1, \underline{a} - \bar{c}M, \bar{b}))\psi = \sigma\psi \quad \text{in } \Omega$$

$$\psi = 0 \quad \text{on } \partial\Omega$$

is positive. In that case, $v \geq \rho$ for large t. With a bit more work, ρ can be estimated from below in terms of the eigenvalues and eigenfunctions of simpler related problems.

It is clear from the above discussion that eigenvalues play a crucial role in the analysis of reaction–diffusion models. Related ideas may also be applied to matrix models. A key result in matrix theory is the Perron–Frobenius theorem, which ensures the existence of a real principal eigenvalue for any nonnegative irreducible matrix. The corresponding infinite dimensional result is the Krein–Rutman theorem, which together with the maximum principle and some technical results in the theory of partial differential equations ensures the existence of a principal eigenvalue for problems of the form

$$\nabla \cdot D(x)\nabla\psi + \vec{B}(x) \cdot \nabla\psi + C(x)\psi = \sigma\psi \quad \text{in } \Omega \tag{6}$$

$$\psi = 0 \quad \text{on } \partial\Omega$$

even though the differential operator is not self–adjoint nor even in divergence form. There are corresponding results for time periodic parabolic equations. For a discussion, see Hess (1991). Use of the theory of eigenvalue problems such as (6) permits the application of the ideas of compressivity, permanence/ uniform persistence, and practical persistence to problems with periodic coefficients and/ or transport mechanisms more complicated than simple diffusion.

References

R.S. Cantrell and C. Cosner, preprint. Practical persistence in ecological models via comparison methods.

R.S. Cantrell, C. Cosner, and V. Hutson, 1991. Permanence in ecological systems with spatial heterogeneity. *Proc. Roy. Soc Edinburgh*, **123A**, 533–559.

R.S. Cantrell, C. Cosner, and V. Hutson, 1993. Permanence in some diffusive Lotka–Volterra models for three interacting species. *Dynamical Systems and Applications*, **2**, 505–530.

Y. Cao and T. Gard, 1993. Uniform persistence for population models with time delay using multiple Lyapunov functions. *Differential and Integral Equations*, **6**, 883–898.

C. Cosner and A.C. Lazer, 1984. Stable coexistence states in the Volterra–Lotka competition model with diffusion. *SIAM Jour. in Appl. Math.*, **44**, 1112–1132.

P. Hess, 1991. *Periodic–Parabolic Boundary Value Problems and Positivity*, Pitman Research Notes in Mathematics #247, Longman Technical and Scientific, Harlow, Essex.

V. Hutson and K. Schmitt, 1992. Permanence and the dynamics of biological systems, *Math. Biosci.*, **111**, 1–71.

H. Smith, 1988. Systems of ordinary differential equations which generate an order preserving flow. A survey of results. *SIAM Rev.*, **30**, 87–113.

Instabilities in Two-Layer Channel Flows

Yuriko Yamamuro Renardy
Department of Mathematics
460 McBryde Hall
Virginia Polytechnic Institute and State University
Blacksburg, Virginia   24061-0123
U.S.A.
email renardyy@math.vt.edu

Abstract
Two-fluid flows exhibit phenomena which form a subject rich in interdisciplinary science. Industrial applications include the formation of bicomponent fibers such as nylons, the production of photographic films and the pipeline transport of crude oil. Theoretically, there are often many possibilities for the shape of the boundary separating the two fluids. On the other hand, experiments show that some shapes are preferred over others. The question thus arises as to which interface shapes can be observed, and what physical processes govern the selection of interface shapes. A first step towards the answer concerns the stability of the interface, and the appearance of new arrangements from unstable ones.

The equations governing flows composed of two immiscible fluids with different viscosities and densities are the Navier-Stokes equations and incompressibility. The consideration of viscoelastic effects will require the replacement of the Navier-Stokes equations with a momentum equation plus a constitutive model. The volumes of the fluids are prescribed, and at the unknown interface position, we have the kinematic free surface condition, the continuity of velocity and shear stress, and the jump in the normal stress is balanced by surface tension. Appropriate boundary conditions are imposed to complete this formulation, e.g., the no-slip condition at solid walls, periodicity in the unbounded direction. There are at least five dimensionless parameters: a Reynolds number, a surface tension parameter, volume ratio, viscosity ratio and density ratio.

The lecture focuses on the stability of two-layer plane channel flow, which has been instructive in understanding phenomena that occur in more complicated flows. Computational and asymptotic results on linear stability will be discussed. We proceed from the linear stability analysis to the study of nonlinear bifurcations. Bifurcation theory proceeds basically in two steps. First, one must show that in the neighborhood of a point where an instability arises, the problem of solving the differential equation can be reduced to a finite dimensional problem. The center manifold theorem is used in this respect. Secondly, the relevant coefficients in the finite dimensional equations must be calculated and the resulting equations must be solved. For shearing flows, we have investigated the stability of the bifurcating travelling wave solutions for plane Couette-Poiseuille flow. Periodicity is prescribed in the horizontal direction. Numerical results on whether the bifurcation is supercritical or subcritical have been obtained for several flow profiles.

The stability of the primary traveling wave solution to sideband perturbations has been examined in various contexts, such as water waves. The usual idea is to derive an amplitude evolution equation known as the Ginzburg-Landau equation. This equation involves an amplitude factor for the critical mode which is allowed to vary slowly as a

function of rescaled space and time variables. The situation with two fluids presents a complication to the usual framework of the Ginzburg-Landau equation. The reason is that in addition to the critical mode leading to traveling waves, there is another neutral mode at zero wavenumber, which corresponds to a shift of the interface. We will incorporate the evolution of this long-wave (or mean-flow) mode, which has been neglected in past work on two-layer flows. If slow modulations are allowed, we obtain, instead of the Ginzburg-Landau equation, a coupled set of three partial differential equations for the amplitudes (of the traveling wave mode, the long-wave mode, and the pressure mode). This raises the possibility that as a result of an instability of the primary traveling wave solution to sideband perturbations, the resulting dynamics may be dominated by the mean-flow mode.

Dispersal-Driven Instability
in a System of Predator-Prey
Integrodifference Equations

Michael Neubert*

October 4, 1994

Abstract

I investigate the dispersal-driven instabilities that arise in a discrete-time predator-prey model formulated as a system of *integrodifference equations*. Integrodifference equations contain two components: (1) difference equations, which model growth and interactions during a sedentary stage and (2) redistribution kernels, which characterize the distribution of dispersal distances that arise during a vagile stage. Redistribution kernels have been measured for a tremendous number of organisms. Using a particular predator-prey model as an example, I show how dispersal-driven instabilities arise and demonstrate the use of a Galerkin method to approximate the stable, spatially patterned solutions which result.

1 Introduction

There is nothing surprising in the observation that organisms track spatial and environmental variation. Indeed, there is an extensive literature on habitat selection (Fretwell and Lucas, 1969; Morse, 1980; Cody, 1985; Bazzaz, 1991; Morrison et al., 1992; Block and Brennan, 1993) and on the role of habitat selection in determining the dispersion of organisms. There is also an older literature on autecological components of niche (Grinell, 1917, 1924, 1928; James et al., 1984), species-specific responses to gradients (Gleason, 1926), and populations (Andrewartha and Birch, 1954, 1984) that argues that environmental heterogeneity is frequently the principal determinant of distribution and abundance.

Despite the ubiquity of environmental heterogeneity, patterned spatial distributions may also arise in *homogeneous* environments (Steele, 1974, 1976, 1978; Mackas and Boyd, 1979; Levin, 1992). Thus it is essential that we ask whether patterns can arise solely as the result of trophic interactions and dispersal. A number of scientists have investigated this question using continuous-time growth models with simple (Fickian) diffusion. The diffusion in these reaction-diffusion models is generally thought of as a stabilizing influence (McMurtrie, 1978), one that homogenizes populations and moderates temporal fluctuations. However, Turing (1952) demonstrated that diffusion may also combine with intra- and interspecific interactions to yield instability and spatial patterns. Turing's concerns were largely developmental. However, his ideas regarding diffusive instability were soon transferred to ecology by Segel and Jackson (1972), Levin and Segel (1976), and Segel and Levin (1976). There is now an extensive literature on diffusive instability and pattern formation in ecology. Okubo (1980), Conway (1984), and Murray (1989) provide useful introductions to this literature.

---

*Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543

A difficulty with reaction-diffusion equations is that they are inappropriate for the innumerable species with discrete, nonoverlapping generations. To circumvent this difficulty, I will consider discrete-time models built around *contact distributions* (Mollison, 1977)—probability distributions for the distance that an organism moves. Discrete-time spatial contact models have a surprisingly long history. They are at the heart of the *problem of random flights* (Markoff, 1912; Chandrasekhar, 1943) wherein a particle undergoes a sequence of independent and random displacements of given distribution. Later, Slatkin (1973), Weinberger (1978, 1984) and Lui (1982a, 1982b, 1983, 1985, 1986, 1989a, 1989b) used them to describe changes in gene frequency. And recently, they have appeared in population ecology as integrodifference equations (Kot and Schaffer, 1986; Hardin et al., 1988a, 1988b, 1990; Kot, 1989, 1992; Anderson, 1991; Hastings and Higgins, 1994; Neubert et al., in press) for populations with discrete nonoverlapping generations and well-defined growth and dispersal stages.

In Section 2 I briefly formulate a system of integrodifference equations for a predator and its prey. Each individual integrodifference equation is built around a redistribution kernel (contact distribution); some possibilities are outlined. Under certain assumptions, the system I study admits a spatially homogeneous solution at which the two species coexist. In Section 3, I perform a linear stability analysis around this solution, and derive a dispersion relation for the growth rates of perturbations of a given wavelength. For some parameter values, a particular wavenumber will have a positive growth rate. Perturbations with this wavenumber will grow, until a new spatially structured solution is obtained. In Section 4, I present a Galerkin method for approximating these spatially structured solutions. A particular example is used throughout to illustrate the methods. Concluding remarks are relegated to Section 6.

2 Integrodifference Equations

Difference Equations

I wish to consider some simple models for interacting populations—predator and prey or host and parasitoid—that grow, interact, and disperse, in synchrony, on a continuous one-dimensional habitat. Each model will be built on top of a system of first-order difference equations,

$$N_{t+1} = f(N_t, P_t), \tag{1a}$$
$$P_{t+1} = g(N_t, P_t), \tag{1b}$$

for the levels of the two populations, N_t and P_t, at time t. There is a long history of such difference equations (Nicholson, 1933; Nicholson and Bailey, 1935; Hassell, 1978). Some of the better-studied systems are known to exhibit complicated and/or chaotic dynamics (Beddington et al., 1975; Gumowski and Mira, 1980; Lauwerier and Metz, 1986; Hadeler and Gerstmann, 1990; Neubert and Kot, 1992).

A particular example of system (1) is the predator-prey model

$$N_{t+1} = N_t \exp\left[r\left(1 - N_t - P_t\right)\right], \tag{2a}$$
$$P_{t+1} = cP_t N_t, \tag{2b}$$

(Neubert and Kot, 1992). This model exhibits most of the same qualitative behavior as the traditional Lotka-Volterra model

$$N_{t+1} = (1+r)N_t - rN_t^2 - cN_tP_t, \qquad (3a)$$

$$P_{t+1} = cN_tP_t, \qquad (3b)$$

(Maynard Smith, 1968). In particular, both models have three equilibria corresponding to (1) extinction of both species, (2) extinction of the predator and survival of the prey at its carrying capacity, or (3) coexistence of both species. This last equilibrium is stable for some parameter values, and can become unstable by Hopf, transcritical, or subcritical flip bifurcations. Systems (2) and (3) differ in that the former maintains first quadrant invariance whereas the later does not. I will use system (2) as the underlying set of difference equations in the examples presented throughout this report.

Dispersal

System (1) makes no allowance for the dispersion of the organisms. To amend this situation, let $N_t(x)$ and $P_t(x)$ represent each population's density in space at the start of the tth generation. I imagine that change occurs as the composition of two distinct stages. Growth, predation, and reproduction occur during a density-dependent sedentary stage. During this stage, $N_t(x)$ is mapped to $f(x, N_t(x), P_t(x))$ while $P_t(x)$ is mapped to $g(x, N_t(x), P_t(x))$. Explicit spatial dependences in f and g (from here on dropped) reflect clinal (spatially varying, time-independent) variation in the parameters. Movement occurs during the second stage. I describe the details of this movement with a pair of linear integral operators that tally the movement from all y to x. The composition of these two stages yields a coupled system of integrodifference equations,

$$N_{t+1}(x) = \int_\Omega k_1(x, y) f(N_t(y), P_t(y)) \, dy, \qquad (4a)$$

$$P_{t+1}(x) = \int_\Omega k_2(x, y) g(N_t(y), P_t(y)) \, dy, \qquad (4b)$$

for the growth and dispersal of N and P in their one-dimensional domain Ω.

The two functions $k_1(x, y)$ and $k_2(x, y)$ at the heart of system (4) are redistribution kernels. Each kernel describes the dispersal of one of the populations from y about y. The product $k_1(x, y) \, dy$ is the probability that an N individual at x at $t+1$ originated, at time t, from an interval of length dy about y. The product $k_2(x, y) \, dy$ provides the same information for P propagules. The two kernels are constrained to be nonnegative. Each kernel may depend on absolute location or on relative distance. If both kernels depend on relative distance, we may rewrite system (4) in terms of convolution integrals:

$$N_{t+1}(x) = \int_\Omega k_1(x - y) f(N_t(y), P_t(y)) \, dy, \qquad (5a)$$

$$P_{t+1}(x) = \int_\Omega k_2(x - y) g(N_t(y), P_t(y)) \, dy, \qquad (5b)$$

I will restrict my attention to redistribution kernels of this type for the remainder of this report.

There are a number of methods for estimating redistribution kernels from observed data (Southwood, 1978; Silverman, 1986). In point of fact, redistribution kernels have been measured for a tremendous number of organisms, ranging from plant spores to small animals (Dobzhansky and Wright, 1943; Wolfenbarger, 1946, 1959, 1975; Cremer, 1966; Platt and Weiss, 1977; Stapanian and

Smith, 1978; Taylor, 1978; Westelaken and Maun, 1985; Okubo and Levin, 1989; Tang, 1989; Willson, 1993). Observed dispersal curves are frequently characterized as being leptokurtic (Bateman, 1950; Okubo, 1980; Howe and Smallwood, 1982; Howe and Westley, 1986; Wilson, 1992).

Theoretical redistribution kernels can also be derived from first principles. For example, consider a propagule that performs a one-dimensional unbiased random walk (Fickian diffusion) during which there is a constant probability of becoming immobile. After a sufficiently long time, the probability density function for the location of the propagule is given by the leptokurtic Laplace or double-exponential distribution (Broadbent and Kendall, 1953; Williams, 1961; Neubert et al., in press):

$$k(x) = \frac{\beta}{2} e^{-\beta |x|} \tag{6}$$

(see Figure 1).

While many measured dispersal curves have their maxima located at the source, a large number are bimodal, with their maxima at some intermediate distance from the parent (Cremer, 1966; Platt and Weiss, 1977; Stapanian and Smith, 1978; Howe et al., 1985). By coupling radial advection with a constant settling rate one can derive the bimodal double-gamma redistribution kernel (Neubert et al., in press):

$$k(x) = \frac{\alpha^2}{2} |x| e^{-\alpha |x|} \tag{7}$$

(see Figure 1). Okubo and Levin (1989) have derived other dispersal curves with nonzero modes which take into account height dependent wind speeds and vertical eddy diffusivity. The kernels (6) and (7) are simple in form and will be used in the examples below.
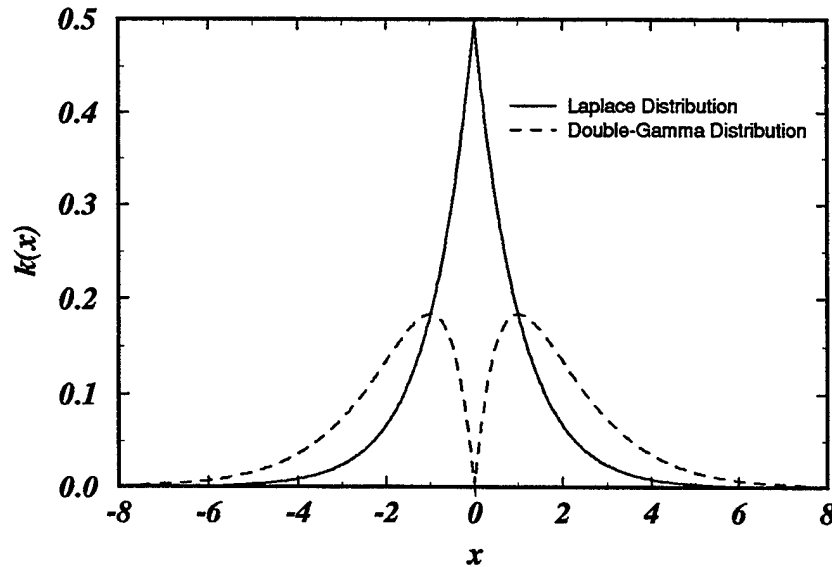
Figure 1: Laplace and double-gamma redistribution kernels. Laplace redistribution kernels result when propagules diffuse and suffer a constant probability of becoming immobile. Double-gamma redistribution kernels are obtained when radial advection is coupled with a constant settling rate. In this figure, $\alpha = \beta = 1$ (see equations [6] and [7]).

183

3 Dispersal-Driven Instabilities

Turing (1952) was the first to appreciate that diffusion may *destabilize* a spatially uniform solution. Diffusive instability occurs when one species diffuses and decays so rapidly that it cannot exert its stabilizing influence upon a second species (Segel and Jackson, 1972). In this section, I study the effects of dispersal on an otherwise stable equilibrium of system (1) and thereby show that dispersal-driven instabilities also occur in systems of integrodifference equations.

Linear Analysis

I start with a pair of difference equations,

$$N_{t+1} = f(N_t, P_t), \tag{8a}$$

$$P_{t+1} = g(N_t, P_t), \tag{8b}$$

with an equilibrium point (N^*, P^*) in the interior of the first quadrant. The equilibrium point must satisfy

$$N^* = f(N^*, P^*), \tag{9a}$$

$$P^* = g(N^*, P^*). \tag{9b}$$

I assume that this equilibrium point is asymptotically stable, with all of the eigenvalues of the Jacobian or community matrix

$$\mathbf{J} = \left(\begin{array}{cc} \frac{\partial f}{\partial N} & \frac{\partial f}{\partial P} \\[2mm] \frac{\partial g}{\partial N} & \frac{\partial g}{\partial P} \end{array} \right) \Bigg|_{(N^*, P^*)} = \left(\begin{array}{cc} a_{11} & a_{12} \\[2mm] a_{21} & a_{22} \end{array} \right) \tag{10}$$

of magnitude less than one.

Next, I add dispersal and consider

$$N_{t+1}(x) = \int_{-\infty}^{\infty} k_1(x-y) \, f(N_t(y), P_t(y)) \, dy, \tag{11a}$$

$$P_{t+1}(x) = \int_{-\infty}^{\infty} k_2(x-y) \, g(N_t(y), P_t(y)) \, dy. \tag{11b}$$

The domain is infinite so that I may start with a spatially uniform steady state. I further limit myself to kernels $k_1(x-y)$ and $k_2(x-y)$ that are symmetric (or even) and nonnegative. Finally, as a matter of convenience, I assume that there is no mortality during dispersal,

$$\int_{-\infty}^{\infty} k_i(x) \, dx = 1, \tag{12}$$

so that (N^*, P^*) is also a spatially uniform steady state of the full integrodifference system (11). Can dispersal destabilize (N^*, P^*)? Consider perturbations of the steady state of the form

$$N_t(x) = N^* + n_t(x), \tag{13a}$$

$$P_t(x) = P^* + p_t(x). \tag{13b}$$

For sufficiently small perturbations, we may linearize about (N^*, P^*),

$$n_{t+1}(x) = \int_{-\infty}^{\infty} k_1(x-y) \, [a_{11} n_t(y) + a_{12} p_t(y)] \, dy, \tag{14a}$$

$$p_{t+1}(x) = \int_{-\infty}^{\infty} k_2(x-y) \, [a_{21} n_t(y) + a_{22} p_t(y)] \, dy. \tag{14b}$$

The a_{ij} are once again given by the Jacobian of the underlying system of difference equations evaluated at (N^*, P^*) (see equation [10]). If the perturbations n_t and p_t decay to zero, (N^*, P^*) is asymptotically stable.

The integrals in equations (14) are convolutions. Since convolution integrals simplify under Fourier transformation, we may use the Fourier transform pair

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} f(x)\,dx, \tag{15a}$$

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega x} \hat{f}(\omega)\,d\omega, \tag{15b}$$

and the transformed perturbations

$$\hat{n}_t(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} n_t(x)\,dx, \tag{16a}$$

$$\hat{p}_t(\omega) = \int_{-\infty}^{\infty} e^{i\omega x} p_t(x)\,dx, \tag{16b}$$

to simplify our analysis. Indeed, upon taking the transform of system (14), I get

$$\begin{pmatrix} \hat{n}_{t+1} \\ \hat{p}_{t+1} \end{pmatrix} = \mathbf{KJ} \begin{pmatrix} \hat{n}_t \\ \hat{p}_t \end{pmatrix}, \tag{17}$$

where $\mathbf{J}$ is the Jacobian or community matrix

$$\mathbf{J} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \tag{18}$$

and $\mathbf{K}$ is given by

$$\mathbf{K} = \begin{pmatrix} \hat{k}_1(\omega) & 0 \\ 0 & \hat{k}_2(\omega) \end{pmatrix}. \tag{19}$$

Decay of the $\hat{n}_t(\omega)$ and $\hat{p}_t(\omega)$ for all ω guarantees the decay of $n_t(x)$ and $p_t(x)$ in $L_2(\mathbb{R})$.

In the absence of dispersal,

$$\begin{pmatrix} n_{t+1} \\ p_{t+1} \end{pmatrix} = \mathbf{J} \begin{pmatrix} n_t \\ p_t \end{pmatrix}. \tag{20}$$

With dispersal,

$$\begin{pmatrix} \hat{n}_{t+1} \\ \hat{p}_{t+1} \end{pmatrix} = \mathbf{KJ} \begin{pmatrix} \hat{n}_t \\ \hat{p}_t \end{pmatrix}. \tag{21}$$

By assumption, (N^*, P^*) is asymptotically stable in the absence of dispersal, with all of the eigenvalues of $\mathbf{J}$ of modulus less than one. Dispersal-driven instability arises if the matrix $\mathbf{KJ}$, in turn, has one or more eigenvalues of modulus greater than one.

One can say more. All eigenvalues of $\mathbf{J}$ have modulus less than one if the trace and determinant of $\mathbf{J}$ satisfy the Jury test (Jury, 1964, 1974)

$$1 - \operatorname{tr}\mathbf{J} + \det\mathbf{J} > 0, \tag{22a}$$

$$1 + \operatorname{tr}\mathbf{J} + \det\mathbf{J} > 0, \tag{22b}$$

$$1 - \det\mathbf{J} > 0. \tag{22c}$$

For dispersal-driven instability, conditions (22) must still hold. In addition, one of the corresponding conditions on **KJ**,

$$1 - \mathrm{tr}\,(\mathbf{KJ}) + \det(\mathbf{KJ}) \;>\; 0, \qquad (23a)$$

$$1 + \mathrm{tr}\,(\mathbf{KJ}) + \det(\mathbf{KJ}) \;>\; 0, \qquad (23b)$$

$$1 - \det(\mathbf{KJ}) \;>\; 0, \qquad (23c)$$

must be violated. The matrix **K** depends on the wavenumber ω; dispersal-driven instability will occur (if it occurs at all) for limited ranges of ω.

Inequality (23a) guarantees that no real eigenvalue of **KJ** is larger than +1. If we violate this inequality, the uniform steady state will lose its stability to a spatially structured solution. I refer to this situation as a plus-one bifurcation. If we reverse inequality (23b), so that an eigenvalue passes below -1, stability is lost to a spatially structured, time-periodic solution of period two (a two-cycle); we have a minus-one bifurcation. The reversal of (23c) would lead to a Hopf bifurcation (with a complex conjugate pair of eigenvalues passing through the unit circle), would that it could occur. However, two-species dispersal-driven Hopf bifurcations are impossible (Kot and Schaffer 1986; Kot 1989).

Dispersion Relation

Inequalities (23) determine whether small amplitude perturbations of a particular wavenumber will eventually grow or decay. They do not, however, give the relative growth rates of each wavenumber. We would like to know, for example, which wavenumber will initially grow the fastest. (At least for the examples presented later, it is this wavenumber which best characterizes the spatially patterned solutions.) We must therefore calculate the growth rates explicitly.

The linear growth rate, $\lambda_{\max}$, for each wave number ω, is given by the largest (in modulus) eigenvalue of the matrix **KJ**. The eigenvalues are in turn given by the solutions of the characteristic equation

$$\lambda^2 - \mathrm{tr}\,(\mathbf{KJ})\,\lambda + \det(\mathbf{KJ}) = 0. \qquad (24)$$

Thus,

$$\lambda(\omega) = \frac{1}{2}\left\{ \mathrm{tr}\,(\mathbf{KJ}) \pm \sqrt{[\mathrm{tr}\,(\mathbf{KJ})]^2 - 4\det(\mathbf{KJ})} \right\}. \qquad (25)$$

The expression $\lambda_{\max} = \lambda_{\max}(\omega)$ is called the *dispersion relation*. If $|\lambda_{\max}|$ is larger than 1 at a particular wavenumber, that wavenumber is unstable. The "most dangerous" wave number, ω^*, is the one which produces the largest $|\lambda_{\max}|$. Figure (2) shows an example of a typical dispersion relation.

Example

Consider as an example system (2), which in the absence of the predator reduces to the Ricker curve (Ricker, 1954). System (2) posseses the unique equilibrium

$$N^* \;=\; \frac{1}{c}, \qquad (26a)$$

$$P^* \;=\; 1 - \frac{1}{c}, \qquad (26b)$$

where predator and prey coexist. The stability of this equilibrium is determined by the Jacobian
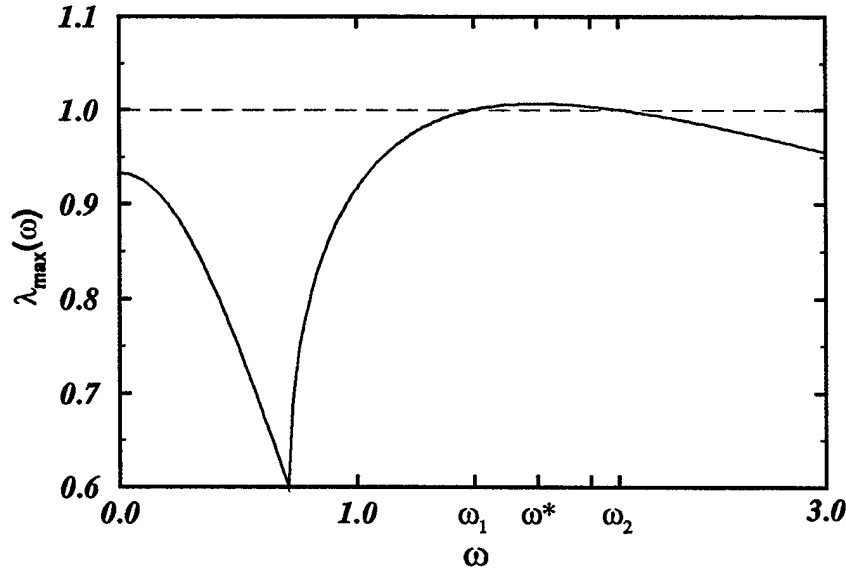
Figure 2: A typical dispersion relation. The linear growth rate $\lambda_{\max}$ is plotted as a function of the wavenumber ω. Any wavenumber between $\omega_1 \approx 1.50$ and $\omega_2 \approx 2.12$ is unstable. $\omega^* \approx 1.77$ is the most unstable wavenumber. In this figure $r = 0.9$, $c = 1.75$, $\alpha = 1.3$ and $\beta = 10.0$.

or community matrix $\mathbf{J}$ evaluated at equilibrium point (26):

$$\mathbf{J} = \begin{pmatrix} 1 - \frac{r}{c} & -\frac{r}{c} \\ c - 1 & 1 \end{pmatrix}. \tag{27}$$

The Jury conditions (22) guarantee that this equilibrium is stable for

$$1 < c < 2, \tag{28a}$$

$$0 < r < \frac{4c}{3 - c}. \tag{28b}$$

If $c < 1$, the predator is not efficient enough to maintain itself and all trajectories are drawn to an equilibrium at which the predator is absent and the prey exist at their carrying capacity. If c becomes larger than two a Hopf bifurcation results, and (except at a few critical values of c) the equilibrium (N^*, P^*) generally yields its stability to a stable invariant circle. Finally, if r becomes larger than $4c/(3 - c)$, a subcritical flip bifurcation ensues (Neubert and Kot, 1992): the predator population undergoes a catastrophic collapse to extinction and the prey population is left to its natural tendencies.

To these dynamics, I added combinations of two redistribution kernels, the double-gamma distribution (7), and the Laplace distribution (6). I thus considered four different models (two species, each with one of two kernels). For a high-r prey, sufficiently high *predator* overdispersal invariably set off a minus-one bifurcation, with pattern formation. In three of the four models (the exception involved two Laplace distributions), sufficiently high *prey* overdispersal led to a plus-one bifurcation, again with pattern formation. Plus-one bifurcations did not require high prey intrinsic rates of growth. Indeed, any point in region (28) could be destabilized by sufficiently high prey overdispersal.

187

Figures 3 and 4 show the outcome of a typical numerical simulations. The parameters r and c were chosen so as to lie in region (28). I added double-gamma-distribution dispersal to the prey and Laplace-distribution dispersal to the predator so as to give

$$N_{t+1}(x) = \int_\Omega k_1(x-y)\, N_t(y) e^{r[1-N_t(y)-P_t(y)]}\, dy, \tag{29a}$$

$$P_{t+1}(x) = \int_\Omega k_2(x-y)\, [cN_t(y)P_t(y)]\, dy, \tag{29b}$$

with

$$k_1(x-y) = \frac{\alpha^2}{2}\, |x-y|\, e^{-\alpha|x-y|}, \tag{30a}$$

$$k_2(x-y) = \frac{\beta}{2} e^{-\beta|x-y|}. \tag{30b}$$

Mathematical analyses were performed on an infinite domain. Numerical simulations were, in contrast, performed on a large finite domain Ω using a fast FFT algorithm (Anderson, 1991). In particular, I iterated system (29) until I obtained a stable solution. Figure 3 shows the resulting steady state for $\alpha = 13$ and $\beta = 5$. The interior of this solution is flat. There is variation toward the ends, but this is an edge effect stemming from the finiteness of the domain. As I increased the variance of the predator's dispersal kernel (by decreasing β), the "homogeneous" solution eventually became unstable. In its place I obtained a spatially structured two-cycle. This period doubling is characteristic of a minus-one bifurcation.

In contrast, when I increased the variance of the prey's redistribution kernel (by decreasing α), I obtained a plus-one bifurcation: the spatially homogeneous solution was replaced by a time-independent, but spatially structured, steady state (see Figure 4). A plot of the dispersion relation for this case is shown in Figure 2. Note that the most unstable wavenumber ($\omega^* \approx 1.77$) correctly predicts the wavelength of the solution ($2\pi/\omega^* \approx 3.55$) far from the boundary.

4   The Galerkin Approximation

The task before us now is to predict the resulting spatially structured solution, given a perturbation with a particular wave number. A Galerkin method suits this purpose.

Equations (11) for the perturbations $n_t(x)$ and $p_t(x)$ may be written as,

$$n_{t+1}(x) = \int_{-\infty}^{\infty} k_1(x-y) \left\{ \sum_{l=1}^{\infty} \frac{1}{l!} \left[n_t(y)\frac{\partial}{\partial N} + p_t(y)\frac{\partial}{\partial P} \right]^l f \right\}\Bigg|_{N^*,P^*} dy \tag{31a}$$

$$p_{t+1}(x) = \int_{-\infty}^{\infty} k_2(x-y) \left\{ \sum_{l=1}^{\infty} \frac{1}{l!} \left[n_t(y)\frac{\partial}{\partial N} + p_t(y)\frac{\partial}{\partial P} \right]^l g \right\}\Bigg|_{N^*,P^*} dy \tag{31b}$$

where I have used the Taylor series expansions of the functions f and g near the equilibrium solution (N^*, P^*). I assume that the perturbations are even and periodic[1] with period L

$$n_t(x+L) = n_t(x), \tag{32a}$$

$$p_t(x+L) = p_t(x). \tag{32b}$$

---

[1]This assumption has the effect of discretizing the spectra of the linear integral operators in (11), which in the absence of (32) have continuous spectra.
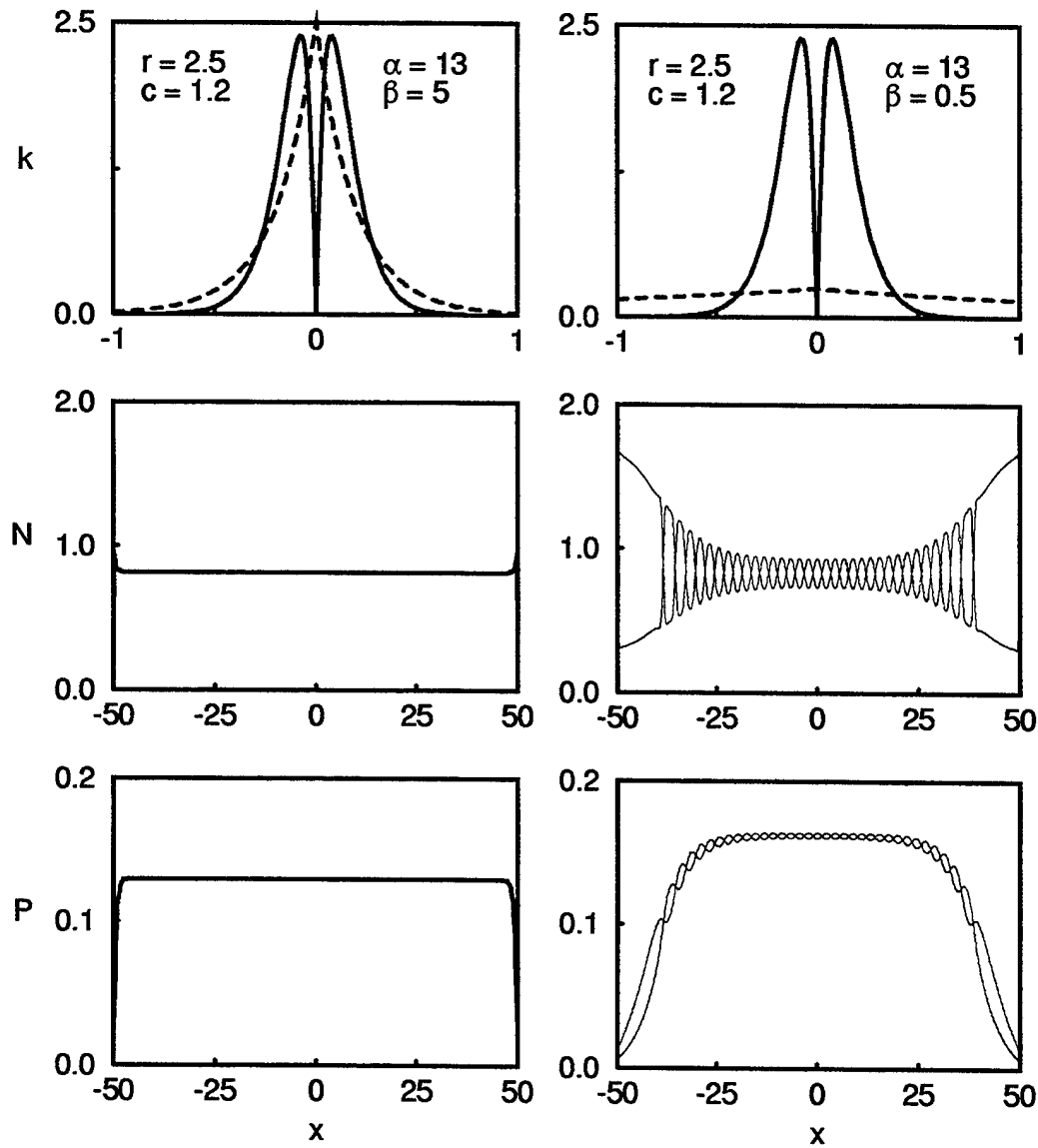
Figure 3: A minus-one bifurcation. For the indicated values of r, c, and α, increasing predator dispersal from $\beta = 5$ to $\beta = 0.5$ destabilizes a spatially homogeneous solution. The resulting instability grows until a stable, spatially-patterned, two-cycle is reached. I iterated system (29) 5,000 times on a domain of length 100 using the initial conditions $N(x) = 0.833$, $P(x) = 0.167$ and 4,096 gridpoints prior to plotting the two-cycle. The top panel illustrates the redistribution kernels for the predator (dashed line) and for the prey (solid line) before (left) and after (right) the bifurcation. The middle and bottom panels show the prey and predator distributions for succeeding iterations, after convergence, before (left) and after (right) the bifurcation.
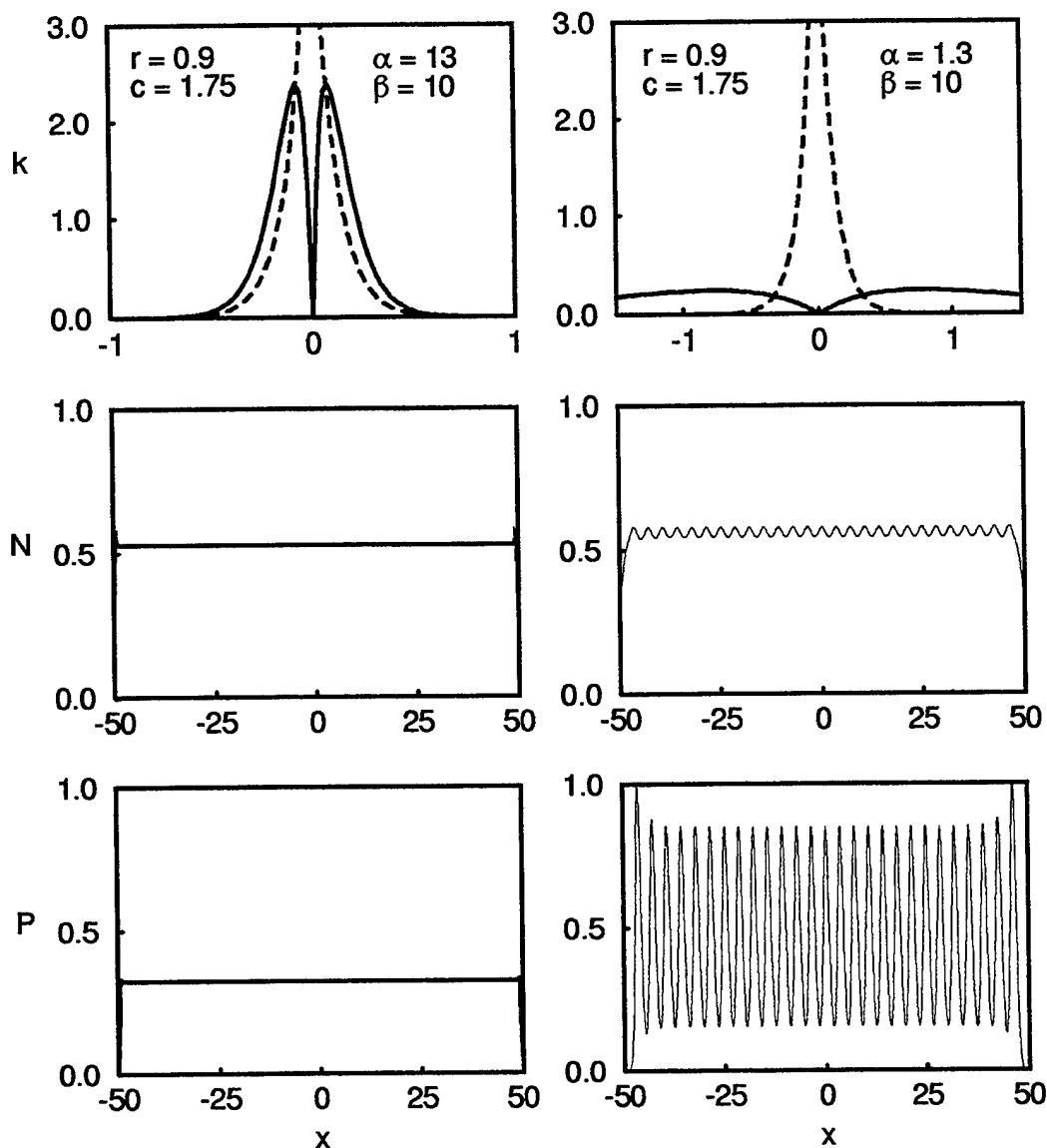
Figure 4: A plus-one bifurcation. For the indicated values of r, c, and β, increasing prey dispersal from $\alpha = 13$ to $\alpha = 1.3$ destabilizes a spatially homogeneous solution. The resulting instability grows until a stable, spatially-patterned, equilibrium is reached. I iterated system (29) 5,000 times on a domain of length 100 using the initial conditions $N(x) = 0.571$, $P(x) = 0.429$ and 4,096 grid points prior to plotting the new equilibrium. The layout of this figure mirrors that of Figure 3.

(The analysis for general periodic functions is similar, but more tedious.) I can then approximate the perturbations as truncated cosine series with time-dependent coefficients:

$$n_t(x) = \sum_{j=0}^{M} a_j(t) \cos \omega_j x, \qquad (33a)$$

$$p_t(x) = \sum_{j=0}^{M} b_j(t) \cos \omega_j x, \qquad (33b)$$

where

$$\omega_j \equiv \frac{2\pi j}{L}. \qquad (34)$$

I then substitute the expressions (33) back into equations (31) to obtain difference equations for the coefficients a_i and b_i. I illustrate this procedure next.

Before proceeding, it is useful to introduce some notation. Define the linear operators which determine the coefficients in the linear, quadratic, and cubic parts of the Taylor series exansion as

$$\mathbf{L}_i \equiv a_i \partial_N + b_i \partial_P, \qquad (35a)$$

$$\mathbf{Q}_{i,j} \equiv a_i a_j \partial_{NN}^2 + 2 a_i b_j \partial_{NP}^2 + b_i b_j \partial_{PP}^2, \qquad (35b)$$

$$\mathbf{C}_{i,j,k} \equiv a_i a_j a_k \partial_{NNN}^3 + 3 a_i a_j b_k \partial_{NNP}^3 + 3 a_i b_j b_k \partial_{NPP}^3 + b_i b_j b_k \partial_{PPP}^3. \qquad (35c)$$

Where, for example, the symbol $\partial_{NP}^2 = \frac{\partial^2}{\partial N \partial P}$. Using expressions (33) and (35) in equations (31) I obtain

$$
\sum_{n=0}^{M} a_n(t+1) \cos \omega_n x = \int_{-\infty}^{\infty} k_1(x-y) \Bigg\{ \sum_{j=0}^{M} \mathbf{L}_j f(N^*, P^*) \cos \omega_j y
$$
$$
+ \frac{1}{2} \sum_{j=0}^{M} \sum_{k=0}^{M} \mathbf{Q}_{j,k} f(N^*, P^*) \cos \omega_j y \, \cos \omega_k y
$$
$$
+ \frac{1}{6} \sum_{i=0}^{M} \sum_{j=0}^{M} \sum_{k=0}^{M} \mathbf{C}_{i,j,k} f(N^*, P^*) \cos \omega_i y \, \cos \omega_j y \, \cos \omega_k y
$$
$$
+ \cdots \Bigg\} dy, \qquad (36a)
$$

$$
\sum_{n=0}^{M} b_n(t+1) \cos \omega_n x = \int_{-\infty}^{\infty} k_2(x-y) \Bigg\{ \sum_{j=0}^{M} \mathbf{L}_j \, g(N^*, P^*) \cos \omega_j y
$$
$$
+ \frac{1}{2} \sum_{j=0}^{M} \sum_{k=0}^{M} \mathbf{Q}_{j,k} \, g(N^*, P^*) \cos \omega_j y \, \cos \omega_k y
$$
$$
+ \frac{1}{6} \sum_{i=0}^{M} \sum_{j=0}^{M} \sum_{k=0}^{M} \mathbf{C}_{i,j,k} \, g(N^*, P^*) \cos \omega_i y \, \cos \omega_j y \, \cos \omega_k y
$$
$$
+ \cdots \Bigg\} dy. \qquad (36b)
$$

I can simplify equations (36) using the identity

$$\cos \omega_j \cos \omega_k = \frac{1}{2} \left[\cos(\omega_j + \omega_k) + \cos(\omega_j - \omega_k) \right], \qquad (37)$$

and the fact that the k_i are even and have Fourier transforms,

$$\int_{-\infty}^{\infty} k_i(x-y) \cos \omega_j y \, dy = \tilde{k}_i(\omega_j) \cos \omega_j x. \tag{38}$$

Using formulas (37) and (38) in equation (36) produces,

$$\sum_{n=0}^{M} a_n(t+1) \cos \omega_n x = \sum_{j=0}^{M} \mathbf{L}_j f(N^*, P^*) \, \tilde{k}_1(\omega_j) \, \cos \omega_j x$$

$$+ \frac{1}{4} \sum_{j=0}^{M} \sum_{k=0}^{M} \mathbf{Q}_{j,k} f(N^*, P^*) \left\{ \tilde{k}_1(\omega_j + \omega_k) \cos\left[(\omega_j + \omega_k)x\right] + \tilde{k}_1(\omega_j - \omega_k) \cos\left[(\omega_j - \omega_k)x\right] \right\}$$

$$+ \frac{1}{24} \sum_{i=0}^{M} \sum_{j=0}^{M} \sum_{k=0}^{M} \mathbf{C}_{i,j,k} f(N^*, P^*) \Big\{ \tilde{k}_1(\omega_i + \omega_j + \omega_k) \cos\left[(\omega_i + \omega_j + \omega_k)x\right] +$$

$$\tilde{k}_1(\omega_i + \omega_j - \omega_k) \cos\left[(\omega_i + \omega_j - \omega_k)x\right] +$$

$$\tilde{k}_1(\omega_i - \omega_j + \omega_k) \cos\left[(\omega_i - \omega_j + \omega_k)x\right] +$$

$$\tilde{k}_1(\omega_i - \omega_j - \omega_k) \cos\left[(\omega_i - \omega_j - \omega_k)x\right] \Big\} + \cdots, \tag{39a}$$

$$\sum_{n=0}^{M} b_n(t+1) \cos \omega_n x = \sum_{j=0}^{M} \mathbf{L}_j g(N^*, P^*) \, \tilde{k}_2(\omega_j) \, \cos \omega_j x$$

$$+ \frac{1}{4} \sum_{j=0}^{M} \sum_{k=0}^{M} \mathbf{Q}_{j,k} g(N^*, P^*) \left\{ \tilde{k}_2(\omega_j + \omega_k) \cos\left[(\omega_j + \omega_k)x\right] + \tilde{k}_2(\omega_j - \omega_k) \cos\left[(\omega_j - \omega_k)x\right] \right\}$$

$$+ \frac{1}{24} \sum_{i=0}^{M} \sum_{j=0}^{M} \sum_{k=0}^{M} \mathbf{C}_{i,j,k} g(N^*, P^*) \Big\{ \tilde{k}_2(\omega_i + \omega_j + \omega_k) \cos\left[(\omega_i + \omega_j + \omega_k)x\right] +$$

$$\tilde{k}_2(\omega_i + \omega_j - \omega_k) \cos\left[(\omega_i + \omega_j - \omega_k)x\right] +$$

$$\tilde{k}_2(\omega_i - \omega_j + \omega_k) \cos\left[(\omega_i - \omega_j + \omega_k)x\right] +$$

$$\tilde{k}_2(\omega_i - \omega_j - \omega_k) \cos\left[(\omega_i - \omega_j - \omega_k)x\right] \Big\} + \cdots. \tag{39b}$$

I now have two equations for the $2(M+1)$ unknowns $a_n(t)$, $b_n(t)$. The paucity of equations is remedied by the fact that the cosines are mutually orthogonal:

$$\frac{2}{L} \int_0^L \cos \omega_j x \, \cos \omega_k x \, dx = \begin{cases} 0 & \text{if} \quad j \neq k \\ 1 & \text{if} \quad j = k \neq 0 \\ 2 & \text{if} \quad j = k = 0 \end{cases} \tag{40}$$

Multiplying each of equations (39) by $\cos \omega_n x$ and integrating each over the interval $0 < x < L$, I obtain as many equations as unknowns:

$$a_0(t+1) = \left\{ \mathbf{L}_0 + \frac{1}{4} \mathbf{Q}_{0,0} + \frac{1}{24} \mathbf{C}_{0,0,0} + \frac{1}{4} \sum_{j=0}^{M} \mathbf{Q}_{j,j} + \right.$$

$$\left. \frac{1}{24} \sum_{j=0}^{M} \left[\sum_{k=0}^{M-j} \mathbf{C}_{j,k,j+k} + \sum_{k=j}^{M} \mathbf{C}_{j,k,k-j} + \sum_{k=0}^{j} \mathbf{C}_{j,k,j-k} \right] \right\} f(N^*, P^*), \tag{41a}$$

$$a_n(t+1) = \tilde{k}_1(\omega_n)\Biggl\{ \mathbf{L}_n + \frac{1}{4}\left[\sum_{j=0}^{n}\mathbf{Q}_{j,n-j} + \sum_{j=n}^{M}\mathbf{Q}_{j,j-n} + \sum_{j=0}^{M-n}\mathbf{Q}_{j,j+n}\right]$$

$$+ \frac{1}{24}\left[\sum_{j=0}^{n}\sum_{k=0}^{n-j}\mathbf{C}_{j,k,n-j-k} + \sum_{j=0}^{n}\sum_{k=n-j}^{M}\mathbf{C}_{j,k,j+k-n}\right.$$

$$+ \sum_{j=n+1}^{M}\sum_{k=0}^{M+n-j}\mathbf{C}_{j,k,j+k-n} + \sum_{j=0}^{M-n}\sum_{k=0}^{M-n-j}\mathbf{C}_{j,k,j+k+n}$$

$$+ \sum_{j=0}^{n}\sum_{k=0}^{M+j-n}\mathbf{C}_{j,k,n-j+k} + \sum_{j=n+1}^{M}\sum_{k=j-n}^{M}\mathbf{C}_{j,k,n-j+k}$$

$$+ \sum_{j=0}^{M-n}\sum_{k=j+n}^{M}\mathbf{C}_{j,k,k-j-n} + \sum_{j=n}^{M}\sum_{k=0}^{j-n}\mathbf{C}_{j,k,j-k-n}$$

$$\left.+ \sum_{j=M-n+1}^{M}\sum_{k=j+n-M}^{M}\mathbf{C}_{j,k,j-k+n} + \sum_{j=0}^{M-n}\sum_{k=0}^{j+n}\mathbf{C}_{j,k,j-k+n}\right]\Biggr\}f(N^*,P^*), \quad (41b)$$

$$b_0(t+1) = \Biggl\{ \mathbf{L}_0 + \frac{1}{4}\mathbf{Q}_{0,0} + \frac{1}{24}\mathbf{C}_{0,0,0} + \frac{1}{4}\sum_{j=0}^{M}\mathbf{Q}_{j,j} +$$

$$\frac{1}{24}\sum_{j=0}^{M}\left[\sum_{k=0}^{M-j}\mathbf{C}_{j,k,j+k} + \sum_{k=j}^{M}\mathbf{C}_{j,k,k-j} + \sum_{k=0}^{j}\mathbf{C}_{j,k,j-k}\right]\Biggr\}g(N^*,P^*), \quad (42a)$$

$$b_n(t+1) = \tilde{k}_2(\omega_n)\Biggl\{ \mathbf{L}_n + \frac{1}{4}\left[\sum_{j=0}^{n}\mathbf{Q}_{j,n-j} + \sum_{j=n}^{M}\mathbf{Q}_{j,j-n} + \sum_{j=0}^{M-n}\mathbf{Q}_{j,j+n}\right]$$

$$+ \frac{1}{24}\left[\sum_{j=0}^{n}\sum_{k=0}^{n-j}\mathbf{C}_{j,k,n-j-k} + \sum_{j=0}^{n}\sum_{k=n-j}^{M}\mathbf{C}_{j,k,j+k-n}\right.$$

$$+ \sum_{j=n+1}^{M}\sum_{k=0}^{M+n-j}\mathbf{C}_{j,k,j+k-n} + \sum_{j=0}^{M-n}\sum_{k=0}^{M-n-j}\mathbf{C}_{j,k,j+k+n}$$

$$+ \sum_{j=0}^{n}\sum_{k=0}^{M+j-n}\mathbf{C}_{j,k,n-j+k} + \sum_{j=n+1}^{M}\sum_{k=j-n}^{M}\mathbf{C}_{j,k,n-j+k}$$

$$+ \sum_{j=0}^{M-n}\sum_{k=j+n}^{M}\mathbf{C}_{j,k,k-j-n} + \sum_{j=n}^{M}\sum_{k=0}^{j-n}\mathbf{C}_{j,k,j-k-n}$$

$$\left.+ \sum_{j=M-n+1}^{M}\sum_{k=j+n-M}^{M}\mathbf{C}_{j,k,j-k+n} + \sum_{j=0}^{M-n}\sum_{k=0}^{j+n}\mathbf{C}_{j,k,j-k+n}\right]\Biggr\}g(N^*,P^*). \quad (42b)$$

Although they appear complicated, iterating the difference equations (41–42) is a trivial task compared with the numerical solution of the full set of integrodifference equations (29–30). And while they were derived for problems on an infinite domain, Figures (5) thru (8) illustrate that equations (41–42) also perform well in approximating the solutions far from the boundaries of a finite domain.

In Figure 5 I have plotted the trajectories of system (41–42), using only two modes ($M = 1$), and starting from an initial condition which represents a small perturbation (with the critical
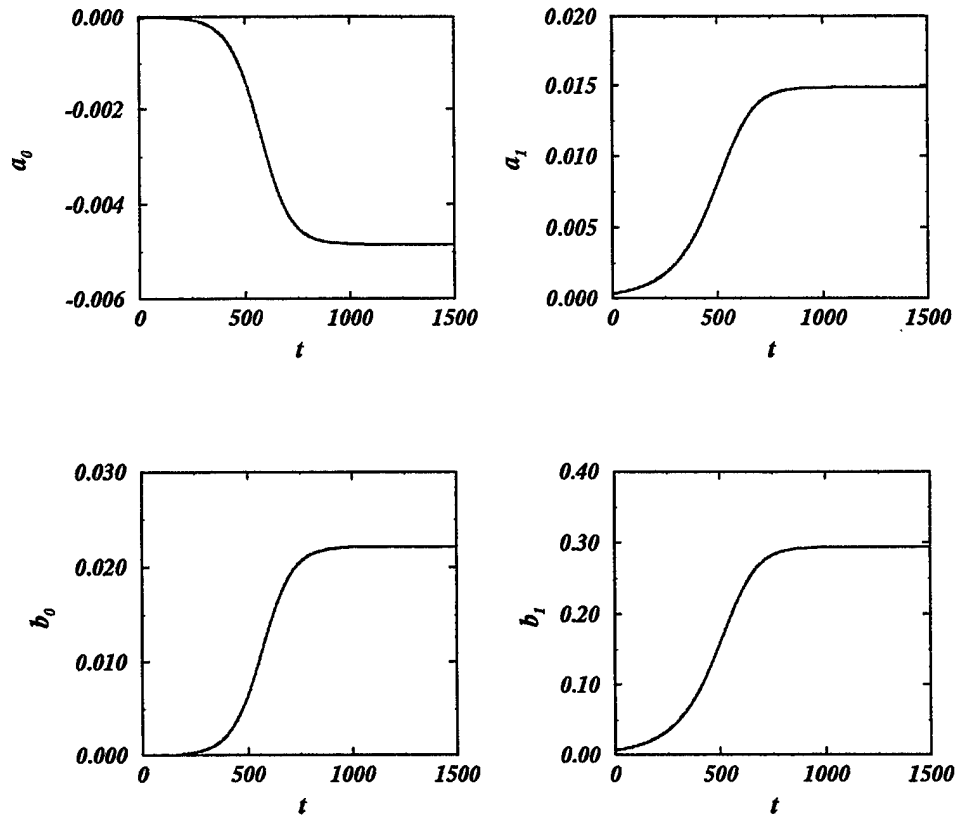
Figure 5: Amplitudes in the Galerkin approximation. Here, I have iterated equations (41-42) 1500 times, with $M = 1$ and $\omega = \omega^*$ from the initial conditions $a_0 = b_0 = b_1 = 0$, $a_1 = 0.001$. The predator-prey interaction is governed by system (29-30), with the parameter values $r = 0.9$, $c = 1.75$, $\alpha = 1.3$, and $\beta = 10.0$. For these parameter values $\omega^* = 1.77152$. Using the asymptotic values of the amplitudes produces a good approximation to the steady-state solutions of system (29-30). See Figure 6.

wavenumber) in the prey density. The parameter values are the same as those used in the right panel of Figure 4. Each of the amplitudes a_i, b_i, monotonically approaches a steady state value and reaches that level after approximately 1000 iterations. Using those asymptotic values in equations (33), I have plotted both the numerically simulated and approximate solutions in Figure 6. The approximation performs reasonably well in the center of the domain, where boundary effects are at a minimum.

I have carried out the same procedure for the minus-one bifurcation of Figure 3. Figure 7 shows the trajectories of system (41–42), using only two modes ($M = 1$), and starting from an initial condition which represents a small perturbation (with the critical wavenumber) in the prey density. The parameter values were the same as those used in the right panel of Figure 3. In this case the amplitudes exhibit growing oscillations until $t \approx 1000$, when a stable two-cycle is acheived. Figure 8 displays both the numerical and approximate solutions, where I have used the asymptotic values of the amplitudes in equations (33). Again the approximation is good only in the middle of the domain.

194
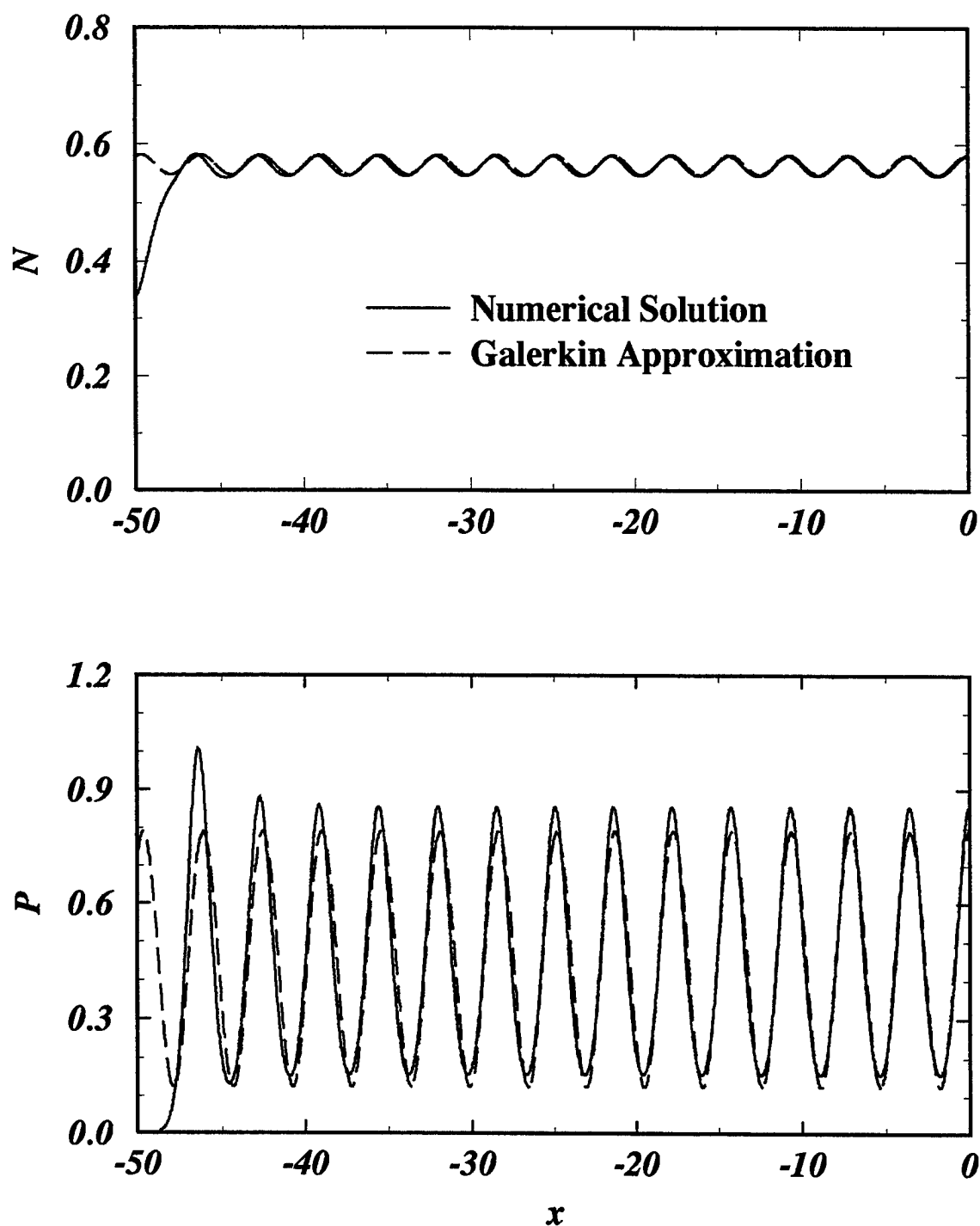
Figure 6: Comparison of approximate and numerical steady-state solutions to system (29-30). The parameter values are as given in Figure 5. Numerical solutions were computed on a domain of length 100. Since the solution is symmetric, only the left half of the domain is shown.

195
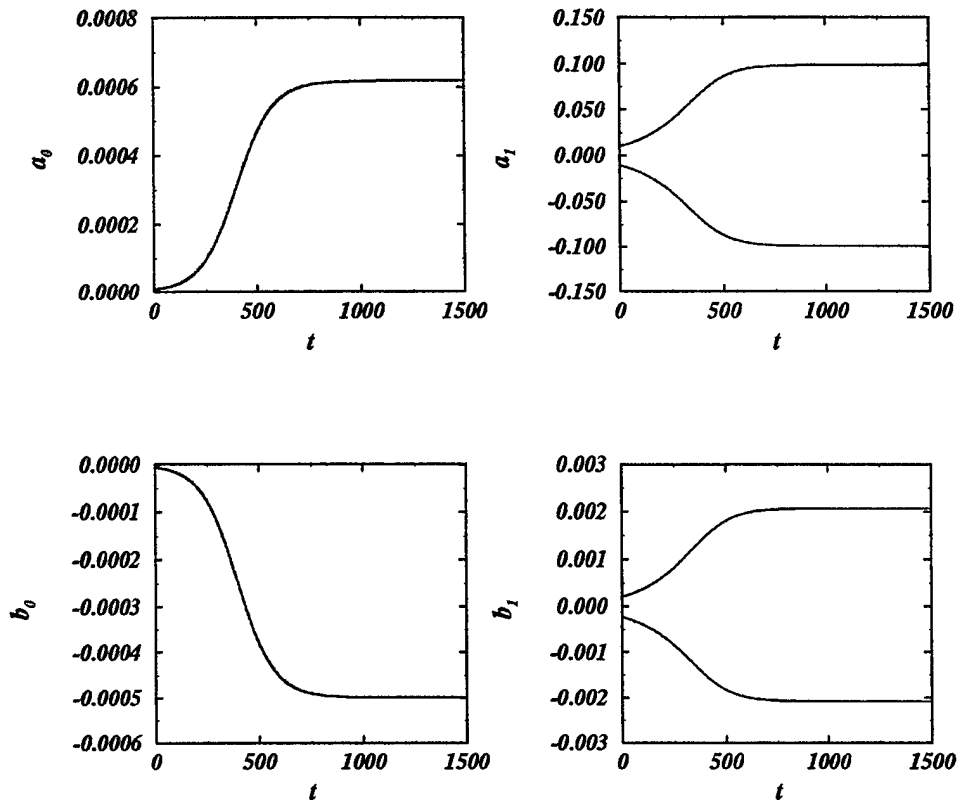
Figure 7: Amplitudes in the Galerkin approximation. Here, I have iterated equations (41-42) 1500 times, with $M = 1$ and $\omega = \omega^*$, from the initial conditions $a_0 = b_0 = b_1 = 0$, $a_1 = 0.001$. The predator-prey interaction is governed by system (29-30), with the parameter values $r = 2.5, c = 1.2$, $\alpha = 13.0$, and $\beta = 0.5$. For these parameter values $\omega^* = 1.37$. The solution alternates between the two branches in the graphs of a_1 and b_1. Using the asymptotic values of the amplitudes produces a good approximation to the two-cycle solutions of system (29-30) in the middle of a large domain. See Figure 8.

5 Discussion

I have shown above how to derive Galerkin-type approximations to the solutions of a predator-prey system of integrodifference equations. Even though the appoximations were derived on an infinite spatial domain, the appear to be good far from the boundary of a large finite domain. With these easily computed approximations in hand, we can answer questions that would be impossible to answer with only a linear analyis and would be computationally expensive to answer by numerical simulation of the full system.

For example, reconsider Figures 5 and 7. They show that (at least for the appropriate parameter values) a plus-one bifurcation decreases the average population size of the prey (a_0) while increasing that of the predator (b_0). Figure 7 shows that the opposite is true for a minus-one bifurcation. One would like to know how the mean values and the amplitudes of the spatial patterns change with changing parameter values.

Figures 9 and 10 show bifurcation diagrams for system (41–42) with $M = 1$. For each value of the bifurcation parameter and for the other parameter values indicated, I calculated the most unstable wavenumber ω^* using Brent's method (Press et al., 1992). System (41–42) was then iterated 2500
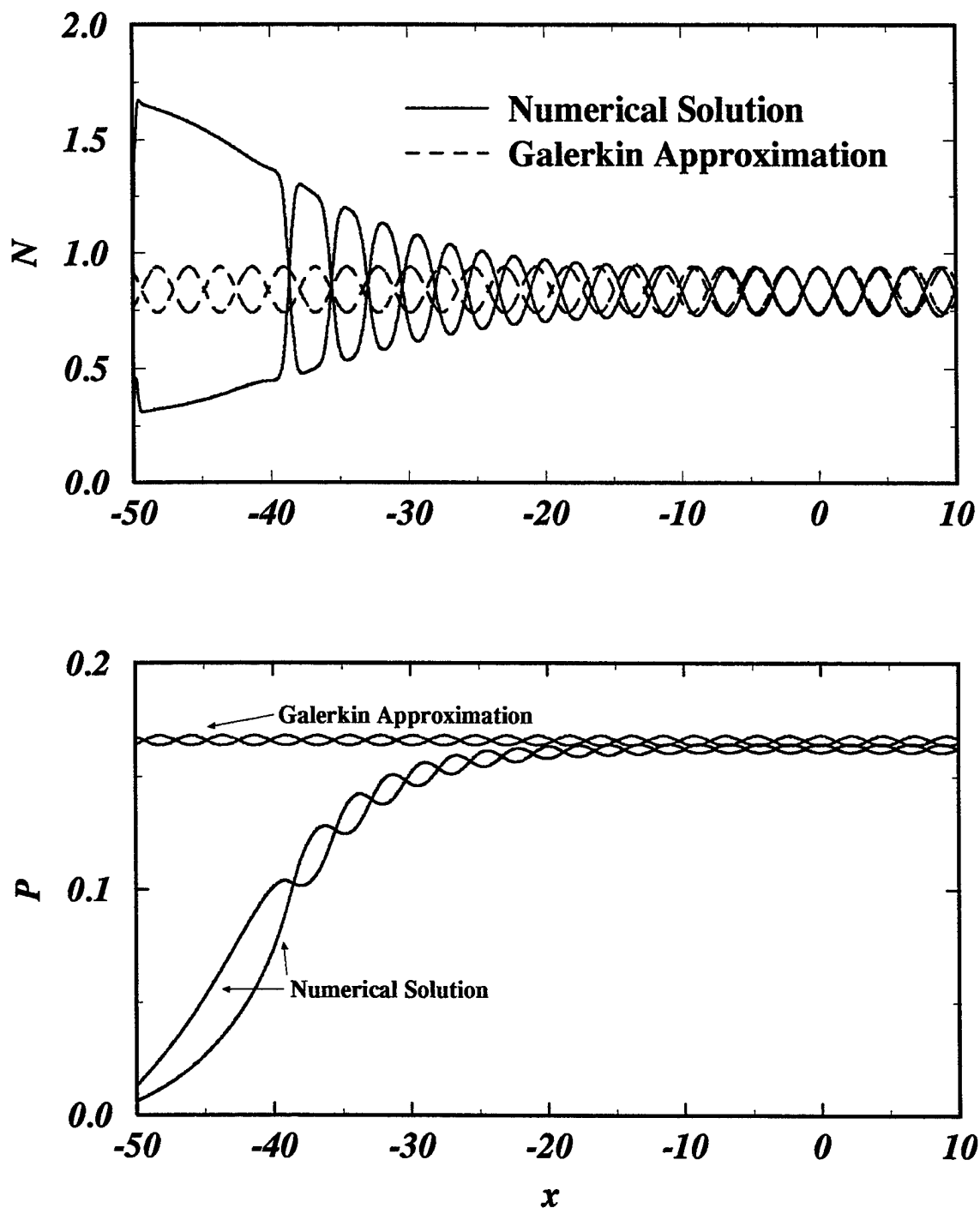
Figure 8: Comparison of approximate and numerical two-cycle solutions to system (29-30). The parameter values are as given in Figure 7. Numerical solutions were computed on a domain of length 100. Since the solution is symmetric, only the left half of the domain is shown. The approximation only becomes good far from the boundaries.

times after which 10 iterations were plotted. Figure 9 displays the plus-one bifurcation that occurs as prey dispersal is increased relative to the predator. After the bifurcation, as β increases, mean prey numbers decrease while mean predator numbers increase. The amplitude of the resulting spatial pattern also increases monotonically with increasing β.

The patterns resulting from the minus-one bifurcation shown in Figure 10 exhibit a more complicated dependence on the dispersal parameters. After the bifurcation, as prey dispersal decreases (for increasing α), mean prey numbers increase to a maximum and then decrease again. Mean predator numbers show the opposite trend, but the extreme value is not obtained for the same level of α. While the amplitude of the spatial pattern in prey numbers increses monotonically, the amplitude of the predator pattern acheives a maximum value for $\alpha \approx 16$ and then decreases.

The linearized stability analysis of Section 3 is not capable of predicting the trends described above. Although they could be computed by numerically iterating the full system, it would be computationally expensive. Another limitation of the linear analyis is its inability to predict secondary bifurcations. Figure 11 shows that the Hopf bifurcation of a spatially homogeneous solution which occurs as c increases beyond 2 (see inequality [28a]) persists even after a plus-one bifurcation. The result, shown in Figure 12, is a spatially patterned solution whose mean values and amplitudes oscillate in time. Although it is practically difficult to compare numerical and approximate solutions for this example, the minimum and maximum amplitudes predicted by the approximation for $c = 2.1$ (see Figure 11) are good approximations to the amplitudes exhibited by the numerical simulations. It would be interesting to see if the other bifurcations of the homogeneous solution predicted by inequalities (28) persist for patterned solutions. It is known, for example, that the subcritical flip bifurcation which occurs for the map (2) does not occur, for some dispersal parameters, for the system of integrodifference equations (4) when both species have Laplace dispersal kernels (Neubert et al., in press). Wether or not the bifurcation occurs on the infinite domain is an open and interesting question.

I have stated throughout this report that the approximation technique used here performs "well" in the interior of a large finite domain. Sources of error in this approximation are at least four-fold: (1) truncation of the Taylor series after cubic terms, (2) truncation of the cosine series after two modes, (3) inclusion of only the most unstable mode even though other nearby modes are also unstable, and (4) effects of a finite domain. (Of course there are also small errors in the numerical simulations.) It would be interesting to quantify each of these sources of error so that one could determine which is the most important.
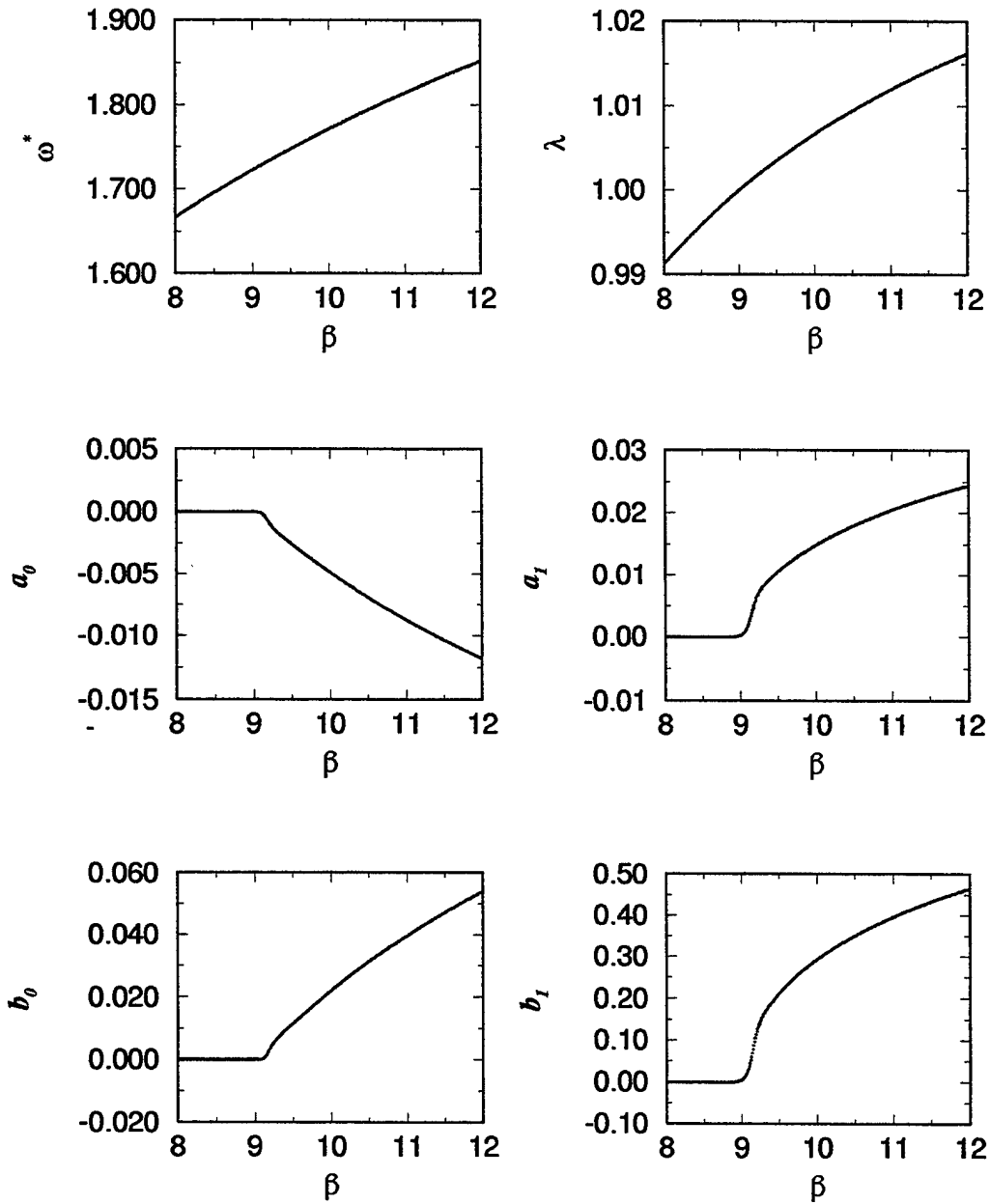
Figure 9: A transcritical bifurcation. For the parameter values $r = 0.9$, $c = 1.75$, $\alpha = 1.3$, and for β ranging between 8 and 12, I calculated the most unstable wavenumber, ω^*, and the linear growth rate of that wavenumber, λ, using Brent's method (Press et al., 1992). For each value of β, I then iterated system (41–42) from the initial condition $a_0 = b_0 = b_1 = 0$ and $a_1 = 0.001$, and plotted 10 iterates after a transient of 2500 iterates. The figure illustrates that the system undergoes a transcritical bifurcation near $\beta = 9.2$ as predator dispersal is diminished relative to the prey.

199
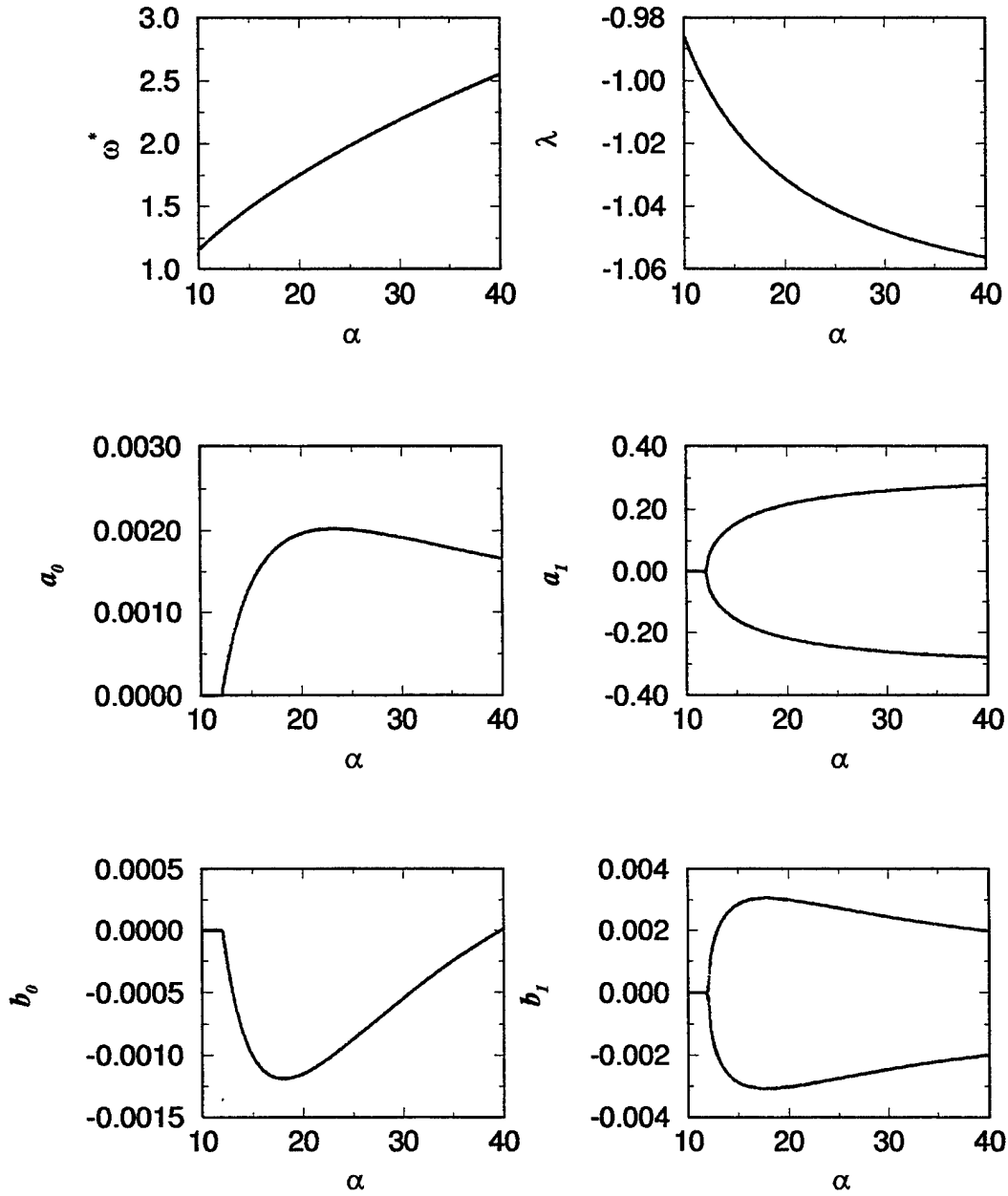
Figure 10: A flip bifurcation. For the parameter values $r = 2.5$, $c = 1.2$, $\beta = 0.5$ and for α ranging between 10 and 40, I calculated the most unstable wavenumber, ω^*, and the linear growth rate of that wavenumber, λ, using Brent's method (Press et al., 1992). For each value of α, I then iterated system (41–42) from the initial condition $a_0 = b_0 = b_1 = 0$ and $a_1 = 0.001$, and plotted 10 iterates after a transient of 2500 iterates. The figure illustrates that the system undergoes a flip bifurcation near $\alpha = 12$ as prey dispersal is diminished relative to the predator.

200
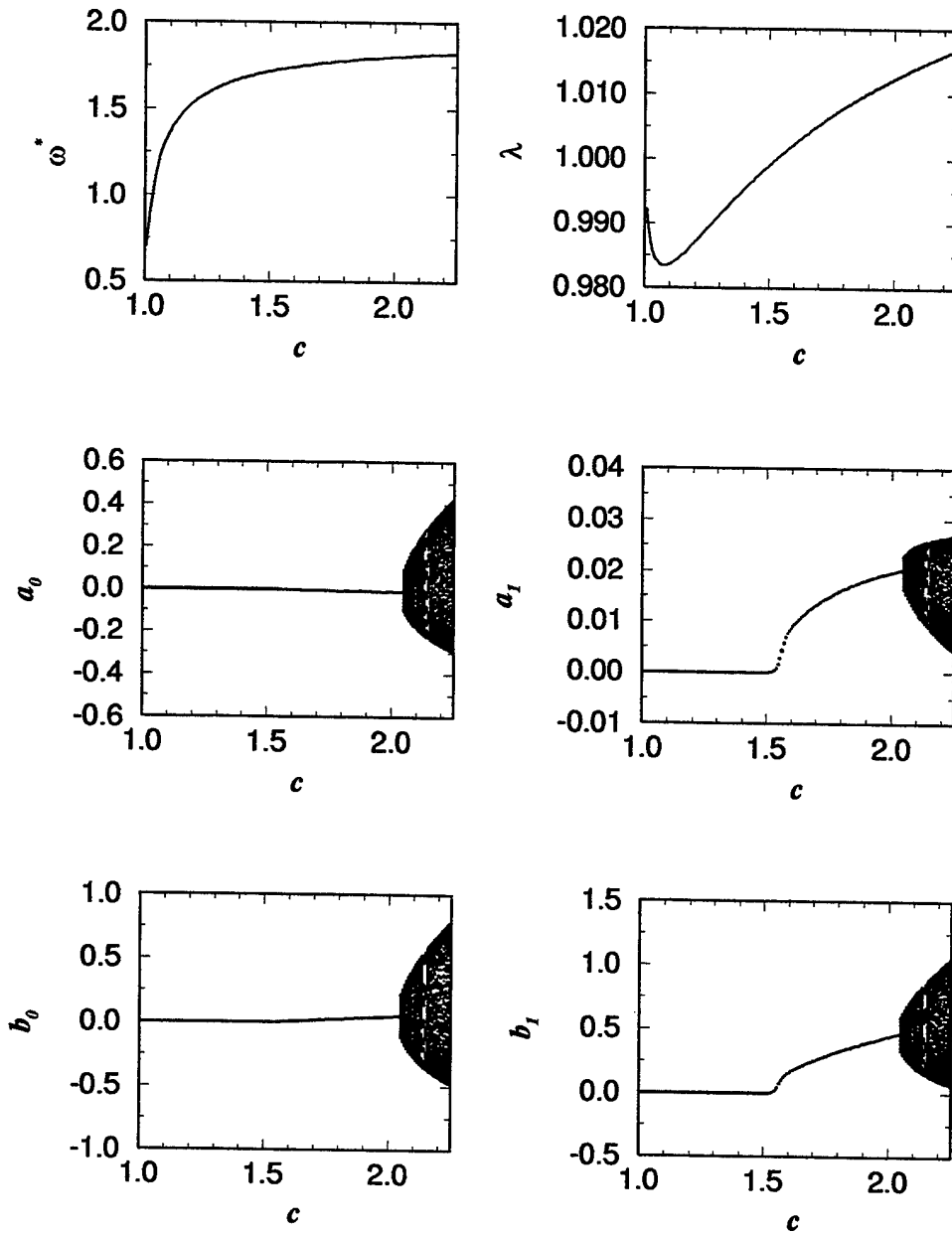
Figure 11: A Hopf bifurcation. For the parameter values $r = 0.9$, $\alpha = 1.3$, $\beta = 10.0$, and for c ranging between 1 and 2.25, I calculated the most unstable wavenumber, ω^*, and the linear growth rate of that wavenumber, λ, using Brent's method (Press et al., 1992). For each value of c, I then iterated system (41–42) from the initial condition $a_0 = b_0 = b_1 = 0$ and $a_1 = 0.001$, and plotted 100 iterates after a transient of 2500 iterates. The figure illustrates that the system first undergoes a transcritical bifurcation near $c = 1.55$. As c becomes larger than two the stable equilibrium undergoes a Hopf bifurcation. See Figure 12.
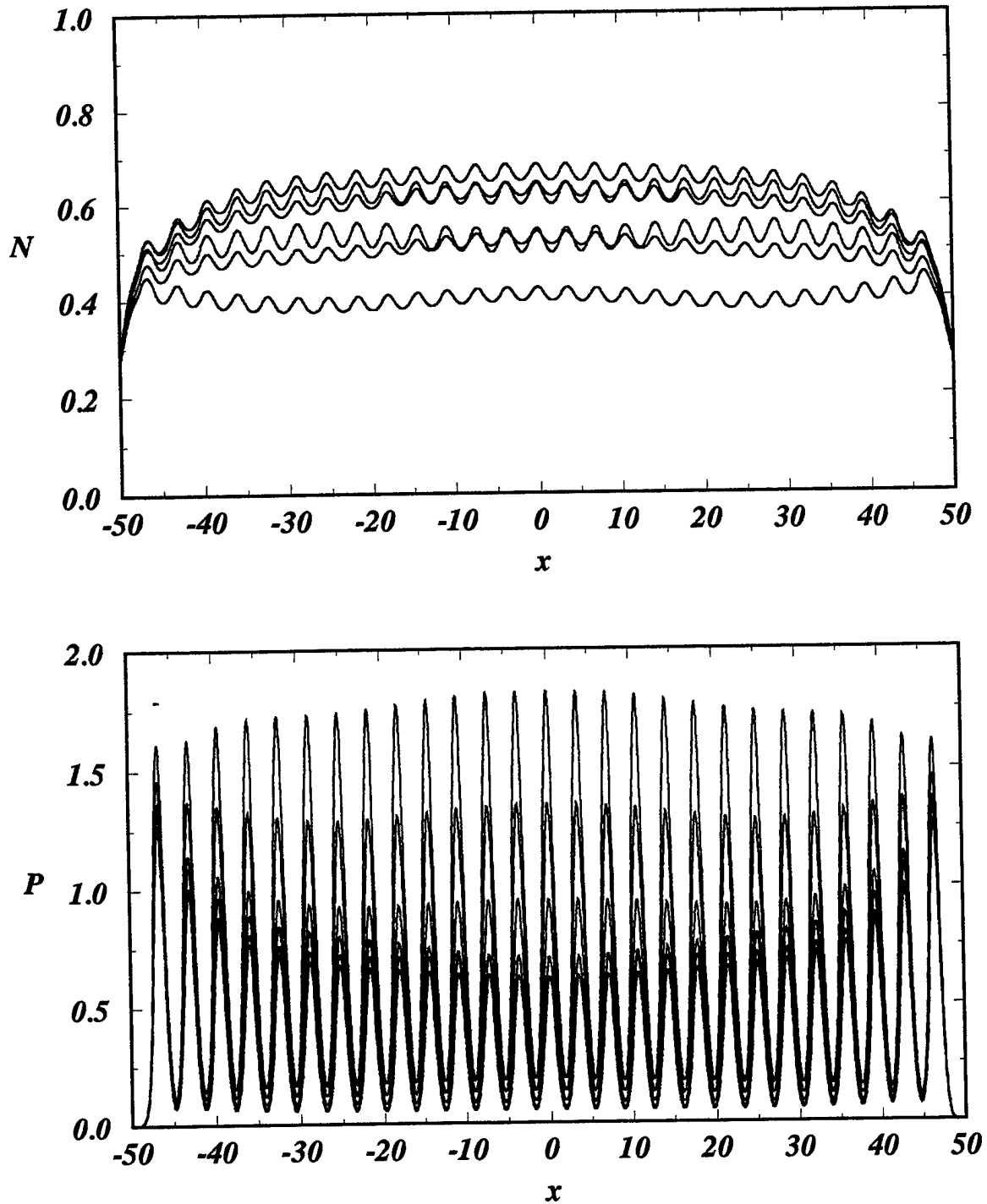
Figure 12: Hopf bifurcation of a spatially structured solution. As shown in the right panel of Figure 4, a steady spatially structured solution is stable for the parameter values $r = 0.9$, $c = 1.75$, $\alpha = 1.3$, and $\beta = 10$. As c is increased, this solution exhibits a Hopf bifurcation (see Figure 11 and equation 28). This figure illustrates 6 iterations of system (29–30) from the initial conditions $N = N^*$, $P = P^*$, using $c = 2.1$.

Acknowledgements

References

Andersen, M. 1991. Properties of some density-dependent integrodifference equation population models. *Mathematical Biosciences* **104**: 135–157.

Andrewartha, H. G. and L. C. Birch. 1954. *The Distribution and Abundance of Animals.* University of Chicago Press, Chicago.

Andrewartha, H. G. and L. C. Birch. 1984. *The Ecological Web: More on the Distribution and Abundance of Animals.* University of Chicago Press, Chicago.

Bateman, A. J. 1950. Is gene dispersion normal? *Heredity* 4:353–363.

Bazzaz, F. A. 1991. Habitat selection in plants. *American Naturalist* **137**:S116–S130.

Beddington, J. R., C. A. Free, and J. H. Lawton. 1975. Dynamic complexity in predator-prey models framed in difference equations. *Nature* **255**:58–60.

Block, W. M. and L. A. Brennan. 1993. The habitat concept in ornithology: theory and applications. *Current Ornighology* **11**:35–91.

Broadbent, S. R. and D. G. Kendall. 1953. The random walk of *Trichostrongylus retortaeformis*. *Biometrika* 9:460–465.

Chandrasekhar, S. 1943. Stochastic problems in physics and astronomy. *Reviews of Modern Physics* **15**:1–91.

Cody, M. L. 1985. *Habitat Selection in Birds.* Academic Press, Orlando.

Conway, E. D. 1984. Diffusion and predator-prey interaction: pattern in closed systems. Pages 85–133 *in* W. E. Fitzgibbon, editor. *Partial Differential Equations and Dynamical Systems.* Pitman, Marshfield.

Cremer, K. W. 1966. Dissemination of seed from *Eucalyptus regnans*. *Australian Forestry* **30**:33–37.

Dobzhansky, T., and S. Wright. 1943. Genetics of natural populations, X. Dispersion rates in *Drosophila pseudoobscura*. *Genetics* **28**:304–340.

Fretwell, S. D. and H. L. Lucas, Jr. 1969. On territorial behavior and other factors influencing habitat distribution in birds. I. Theoretical development. *Acta Biotheoretica* **19**:16–36.

Gleason, H. A. 1926. The individualistic concept of the plant association. *Bulletin of the Torrey Botanical Club* **53**:1–20.

Grinell, J. 1917. The niche-relationships of the California Thrasher. *Auk* **34**:427–433.

Grinell, J. 1924. Geography and evolution. *Ecology* **5**:225–229.

Grinell, J. 1928. Presence and absence of animals. *University of California Chronicle* **30**:429–450.

Gumowski, I. and C. Mira. 1980. *Recurrences and Discrete Dynamic Systems.* Springer-Verlag, New York.

Hadeler, K. P. and I. Gerstmann. 1990. The discrete Rosenzweig model. *Mathematical Biosciences* **98**:49–72.

Hardin, D. P., P. Takac, and G. F. Webb. 1988a. A comparison of dispersal strategies for survival of spatially heterogeneous populations. *SIAM Journal of Applied Mathematics* **48**:1396–1423.

Hardin, D. P., P. Takac, and G. F. Webb. 1988b. Asymptotic properties of a continuous-space discrete-time population model in a random environment. *Journal of Mathematical Biology* **26**:361–374.

Hardin, D. P., P. Takac, and G. F. Webb. 1990. Dispersion population models discrete in time and continuous in space. *Journal of Mathematical Biology* **28**:1–20.

Hassell, M. P. 1978. *The Dynamics of Arthropod Predator-Prey Systems.* Princeton University Press, Princeton.

Hastings, A. and K. Higgins. 1994. Persistence of transients in spatially structured ecological models. *Science* **263**:1133–1136.

Howe, H. F., E. W. Schupp, and L. C. Westley. 1985. Early consequences of seed dispersal for a neotropical tree (*Virola surinamensis*). *Ecology* **66**:781–791.

Howe, H. F., and J. Smallwood. 1982. Ecology of seed dispersal. *Annual Reviews of Ecology and Systematics* **13**:201–228.

Howe, H. F., and L. C. Westley. 1986. Ecology of pollination and seed dispersal. Pages 185–216 *in* M. J. Crawley, editor. *Plant Ecology.* Blackwell Scientific, Oxford, England.

James, F. C., R. F. Johnston, N. O. Wamer, G. J. Niemi, and W. J. Boecklen. 1984. The Grinellian niche of the Wood Thrush. *American Naturalist* **124**:17–30.

Jury, E. I. 1964. *Theory and Application of the Z-Transform Method.* Wiley, New York.

Jury, E. I. 1974. *Inners and Stability of Dynamic Systems.* Wiley, New York.

Kot, M. 1989. Diffusion-driven period-doubling bifurcations. *BioSystems* **22**:279–287.

Kot, M. 1992. Discrete-time travelling waves: ecological examples. *Journal of Mathematical Biology* **30**:413–436.

Kot, M. and W. M. Schaffer. 1986. Discrete-time growth-dispersal models. *Mathematical Biosciences* **80**:109–136.

Lauwerier, H. A. and J. A. Metz. 1986. Hopf bifurcation in host-parasitoid models. *IMA Journal of Mathematics Applied in Medicine and Biology* **3**:191–210.

Levin, S. A. 1992. The problem of pattern and scale in ecology. *Ecology* **73**:1943–1967.

Levin, S. A. and L. A. Segel. 1976. Hypothesis for the origin of planktonic patchiness. *Nature* **259**:659.

Lui, R. 1982a. A nonlinear integral operator arising from a model in population genetics, I. Monotone initial data. *SIAM Journal of Mathematical Analysis* **13**:913–937.

Lui, R. 1982b. A nonlinear intergal operator arising from a model in population genetics, II. Initial data with compact support. *SIAM Journal of Mathematical Analysis* **13**:938–953.

Lui, R. 1983. Existence and stability of travelling wave solutions of a nonlinear integral operator. *Journal of Mathematical Biology* **16**:199–220.

Lui, R. 1985. A nonlinear integral operator arising from a model in population genetics, III. Heterozygote inferior case. *SIAM Journal of Mathematical Analysis* **16**:1180–1206.

Lui, R. 1986. A nonlinear integral operator arising from a model in population genetics, IV. Clines. *SIAM Journal of Mathematical Analyis* **17**:152–168.

Lui, R. 1989a. Biological growth and spread modeled by systems of recursions, I. Mathematical theory. *Mathematical Biosciences* **93**: 269–295.

Lui, R. 1989b. Biological growth and spread modeled by systems of recursions, II. Biological theory. *Mathematical Biosciences* **93**:297–312.

Mackas, D. L. and C. M. Boyd. 1979. Spectral analysis of zooplankton spatial heterogeneity. *Science* **204**:62–64.

McMurtrie, R. 1978. Persistence and stability of single-species and predator-prey systems in spatially heterogeneous environments. *Mathematical Biosciences* **39**:11–51.

Markoff, A. A. 1912. *Wahrscheinlichkeitsrechnung.* Teubner, Leipzig.

Maynard Smith, J. 1968. *Mathematical Ideas in Biology.* Cambridge University Press, Cambridge.

Mollison, D. 1977. Spatial contact models for ecological and epidemic spread. *Journal of the Royal Statistical Society* **39B**:283–326.

Morrison, M. L., B. G. Marcot, and R. W. Mannan. 1992. *Wildlife-Habitat Relationships: Concepts and Applications.* University of Wisconsin Press, Madison.

Morse, D. H. 1980. *Behavioral Mechanisms in Ecology.* Harvard University Press, Cambridge.

Murray, J. D. 1989. *Mathematical Biology.* Springer-Verlag, Berlin.

Neubert, M. G. and M. Kot. 1992. The subcritical collapse of predator populations in discrete-time predator-prey models. *Mathematical Biosciences* 110:45–66.

Neubert, M. G., M. Kot, and M. A. Lewis. In press. Dispersal and pattern formation in a discrete-time predator-prey model. *Theoretical Population Biology.*

Nicholson, A. J. 1933. The balance of animal populations. *Journal of Animal Ecology* 2:132–178.

Nicholson, A. J. and V. A. Bailey. 1935. The balance of animal populations. *Proceedings of the Zoological Society of London*: 551–598.

Okubo, A. 1980. *Diffusion and Ecological Problems: Mathematical Models.* Springer-Verlag, Berlin.

Okubo, A. and S. A. Levin. 1989. A theoretical framework for data analysis of wind dispersal of seeds and pollen. *Ecology* 70:329–338.

Platt, W. J., and I. M. Weiss. 1977. Resource partitioning and competition within guilds of fugitive plants. *American Naturalist* 111:479–513.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. *Numerical Recipes in C.* Cambridge University Press, Cambridge.

Ricker, W. E. 1954. Stock and recruitment. *Journal of the Fisheries Research Board of Canada* 11:559–623.

Segel, L. A. and J. L. Jackson. 1972. Dissipative structure: an explanation and an ecological example. *Journal of Theoretical Biology* 37:545–559.

Segel, L. A. and S. A. Levin. 1976. Application of nonlinear stability theory to the study of the effects of diffusion on predator-prey interactions. Pages 123–152 *in* R. A. Piccirelli, editor. *Topics in Statistical Mechanics and Biophysics: A Memorial to Julius L. Jackson.* American Institute of Physics, New York.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis.* Chapman & Hall, London.

Slatkin, M. 1973. Gene flow and selection in a cline. *Genetics* 75:733–756.

Southwood, T. R. E. 1978. *Ecological Methods.* Chapman & Hall, London.

Stapanian, M. A. and C. C. Smith. 1978. A model for seed scatterhoarding: coevolution of fox squirrels and black walnuts. *Ecology* 59:884–896.

Steele, J. H. 1974. *The Structure of Marine Ecosystems.* Harvard University Press, Cambridge.

Steele, J. H. 1976. Patchiness. Pages 98–115 *in* D. H. Cushing and J. J. Walsh, editors. *The Ecology of the Sea.* W. B. Saunders, Philadelphia.

Steele, J. H. 1978. *Spatial Pattern in Plankton Communities.* Plenum Press, New York.

Tang, W. 1989. Seed dispersal in the cycad *Zamia pumila* in Florida. *Candian Journal of Botany* 67:2066-2070.

Taylor, R. A. J. 1978. The relationship between density and distance of dispersing insects. *Ecological Entomology* 3:63–70.

Turing, A. 1952. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London* B237:37–72.

Weinberger, H. F. 1978. Asymptotic behavior of a model of population genetics. Pages 47–98 *in* J. Chadam, editor. *Nonlinear Partial Differential Equations and Applications.* Lecture Notes in Mathematics 648. Springer-Verlag, New York.

Weinberger, H. F. 1984. Long-time behavior of a class of biological models. Pages 323–352 *in* W. E. Fitzgibbon, editor. *Partial Differential Equations and Dynamical Systems.* Pitman, London.

Westelaken, I. L. and M. A. Maun. 1985. Spatial pattern and seed dispersal of *Lithospermum caroliniense* on Lake Huron sand dunes. *Canadian Journal of Botany* 63:125–132.

Williams, E. J. 1961. The distribution of larvae of randomly moving insects. *Australian Journal of Biological Sciences* **14**:598–604.

Willson, M. F. 1992. The ecology of seed dispersal. Pages 61–85 *in* M. Fenner, editor. *Seeds: The Ecology of Regeneration in Plant Communities*. CAB International, Wallingford.

Willson, M. F. 1993. Dispersal mode, seed shadows, and colonization patterns. *Vegetatio* **107/108**:261–280.

Wolfenbarger, D. O. 1946. Dispersion of small organisms, distance dispersion rates of bacteria, spores, seeds, pollen and insects. Incidence rates of disease and injuries. *American Midland Naturalist* **35**:1–152.

Wolfenbarger, D. O. 1959. Dispersion of small organisms, incidence of viruses and pollen; dispersion of fungus, spores and insects. *Lloydia* **22**:1–106.

Wolfenbarger, D. O. 1975. *Factors Affecting Dispersal Distances of Small Organisms*. Exposition Press, Hicksville.

Effects of nutrient storage and photoadaptation on phytoplankton ecology

Emmanuel Boss

U. of Washington, School of oceanography, WB-10, Seattle 98195.

Introduction

As phytoplankton are advected in the mixed layer the amount of light they encounter changes. In order to adapt to the varied light environment they developed several mechanisms that permit them to continue to grow in very different light regimes. The most common mechanisms are changes in the number of photosynthetic units and/or their size.

A series of observations was performed by J. Mara (Mara, 1978a, Mara, 1978b) to examine the time evolution of the production rate of phytoplankton in different light regimes. The experiment's set-up consisted of a container in which the cells were grown and harvested to keep their density constant. The specie chosen for the experiments was the coastal chain forming diatom *Lauderia borialis*, which has a typical cell size of $40 \mu m$. Two sets of experiments were conducted:

1. The response to a step light forcing (Fig. 1); after the cells were kept for 12 hours in dark, the cells were exposed to a constant light level for 12 hours. Rate of O_2 production was measured as a function of time for cells exposed to different constant light levels.

2. The response to a "daily" and modulated light forcing (Fig. 2); the cells were kept for 12 hours in dark and then exposed for 12 hours to a time-varying light level.

In all of the experiment the following were observed:

1. The integrated photosynthesis over a day (over 10 hours for the step) was almost the same for all cells in light levels above 5% of the maximal value. Modulated and "daily" light exposure with the same integrated light had the same integrated photosynthesis.

2. For the step function forcing, the rate of O_2 production by the cells that were exposed to low light initially rose and then decreased.

3. An afternoon depression in the rate of production was observed in the time-varying light regime.

In a subsequent paper Denman and Mara (1986) explained the results of the step-function forcing experiment and modeled them as a consequence of a cumulative exposure to light which inhibits the rate of production.

There may be, however, another explanation for the above observation. An internal nutrient storage, filled in the dark phase of the experiment and depleted during the light phase can explain the observed decay in production rate. The final asymptotic tail of the decrease in production rate can be explained as a steady state reached in which production rate equals uptake rate. The initial rise of production rate for the cultures exposed to low light can be explained as a result of photoadaptation.

Nutrient storage, a behavior which is observed in large ($O(10 - 100\mu)$) cells, permits phytoplankton to decouple the nutrient uptake from the light conditions, thus maximizing growth. On the short term, nutrients taken at night are assimilated during the day while on the long term some phytoplankton store even more nutrients than they need for their own growth, passing it to their off-spring ('luxury consumption'). The diatoms in Mara's experiments are capable of nutrients storage.

The equal areas under the production rate curves suggest that the light is not the inhibiting factor (if it were, the cells exposed to strong light would have produced less). Moreover, the initial increase in production rate for cells exposed to low light is not consistent with inhibition. Finally, it is hard to explain why phytoplankton that are exposed to 10 percents of the light level at the surface will be inhibited.

In the following section the alternative explanation of nutrient limitation and light adaptation is tested. A model is used to check whether this hypothesis is consistent with the observation.

A model to test the candidate hypothesis

In order to check whether the above hypothesis is consistent with the observations, a numerical model was designed. It consists of phytoplankton growing in a container. The

cell's nutrients are divided into two compartments; incorporated (N_i) and unincorporated (or stored) nutrient (N_s) (Fig. 3). Only the response to a step-function forcing was modeled, as it is easier to make some convenience assumption for this relatively short experiment. The nutrient is not specified but is assumed to be the limiting factor.

The equation describing the mass budget of the nutrients in the different compartments are:

a. Container mass balance:

$$\frac{dN_c}{dt} = (r - G_c)N_i \tag{1}$$

with N_c denoting the concentration the nutrient in the container, r the respiration rate and G_c the nutrient uptake rate by the cells.

b. Storage mass balance:

$$\frac{dN_s}{dt} = (G_c - G_1)N_i - QN_s \tag{2}$$

Where N_s denotes the storage concentration, Q the harvest rate and G_1 the assimilation rate.

c. Incorporated nutrient's mass balance

$$\frac{dN_i}{dt} = (G_1 - r - Q)N_i \tag{3}$$

d. The uptake rate:

$$G_c = Uf(N_s), \ U \equiv \frac{V_c N_c}{K_c + N_c} \tag{4}$$

The uptake rate is assumed to be of the monod type with half saturation and maximal rate K_c and V_c, respectively. $f(N_s)$ is some monotonically decreasing function of the storage concentration.

e. The assimilation rate:

$$G_1 = Min\left(\frac{\mu_m N_s}{K_s + N_s}, \ p_m\left(1 - e^{\left(\frac{\alpha}{p_m}I\right)}\right)\right) \tag{5}$$

G_1 is formulated such that it will represent limitation due to either light or nutrient. The nutrient function is of the Monod type with a half saturation constant K_s and a maximal

209

assimilation rate μ_m. The light function is a similarly saturating function with a maximal assimilation rate p_m and an initial slope α (Platt et al., 1980). Subsequently α will be assumed to be a photoadaptive parameter, changing as the light environment does. There are two formulations for such an adaptation (Cullen and Lewis (1988):

first order kinetics:

$$\frac{d\alpha}{dt} = \frac{1}{\tau}(\hat{\alpha} - \alpha) \tag{6}$$

and logistic kinetics:

$$\frac{d\alpha}{dt} = \frac{\alpha}{\tau}\left(\frac{\hat{\alpha} - \alpha}{\hat{\alpha}}\right) \tag{7}$$

τ is the adaptation rate and $\hat{\alpha}$ is the asymptotic value of α. In this model it is assumed that $\hat{\alpha}$ is determined by the nutrient's availability and is given as the value of α which makes the light function of the assimilation rate equal to the nutrient function in (5).

In order to minimize the number of free parameters some convenience assumptions are made. First it is assumed that the concentration of the nutrient in the container is constant. The second assumption is that the uptake (U) is constant in both day and night with $f(N_s) = 1$. This will be the case if the storage is never filled up.

The model parameters are determined by the details of the experiment as described in Mara (1978a, 1978b).

1. Equations (2)-(3) were reformulated to per/cell units (the units used by Mara) where a cell denotes a certain (constant) amount of incorporated nutrient.

2. The initial number of cells was taken as n(t=0)=20,000/ml.

3. The constant uptake U was assumed to be equal to the O_2 production rate at the end of the experiment (where $\frac{dN_s}{dt} = 0$) and was fixed at 0.008mmol $O_2/hr/cell$.

4. $N_s(t = 0) = 12Un(t = 0)$, as the cells were kept for 12 hours in the dark.

5. The maximal assimilation rate was taken to be equal to 0.08mmol $O_2/hr/cell$, based on the initial production rate of the cells exposed to 100% light in Fig. 1, which were assumed to be nutrient (rather than light) limited. This determined the ratio $\frac{\mu_m}{K_s + N_s(t=0)}$

210

6. The first set of data from Fig. 1 (t=12.5 min) was used to reproduce a P-I curve. p_m and α were fitted to match this curve.

7. $\tau = 8hr$, in accordance with the data.

8. $r = 0.01$, a commonly used respiration coefficient.

9. $Q = 0.0025\ hr^{-1}$ is given.

Results

The model was run using a 4th-order Runga-Kutta integrator. The logistic adaptation gave results that are more consistent with the data (Fig. 4). Taking $r = 0$ and $Q = 0$ did not changed the results noticeably. The following are observed:

a. The model run seemed to capture most of the features observed in the experiments, including the initial increase of production rate for cells exposed to low light level and its subsequent decay.

b. Final N_s value is less than 8% of its initial value (the storage cannot be totally empty when the uptake is non zero).

c. The exponential decay of the production rate did not match for low light level.

Discussion

The hypothesis presented here is consistent with the following observations:

1. In the oceans 70% of the daily photosynthesis at the surface and 58% of the total water-column photosynthesis occurs before noon (Parsons et al., 1984).

2. In laboratory experiments diel oscillations in photosynthesis occur primarily in large cells, and are decoupled from the cell division cycle for diatoms.

3. When starved, the oscillation amplitude decreases (Harding et al., 1981; Prezelin, 1992).

The assumption of constant uptake in the above model needs some clarification. It may be the case if the cells are diffusion limited, a case in which the container concentration of the limiting nutrient is going to be very small (Mara's container was well stirred, but nutrient weren't added during the experiment), or if the cells are reaction limitation, a case

that can be consistent with high nutrient concentration. Two limiting reactions come to mind. Actual nutrient uptake if it is mediated by enzymatic reactions, and dark reaction limitation, i.e. the assimilation is limiting (Jumars, 1994, personal communication). The above model will be suitable only for the case of uptake limitation.

Nutrient storage seems, at first look, to be a strategy that is always advantageous. By storing nutrient at night one can expect to have a good meal during the day. There is, however, a tradeoff; storage, in large quantities, requires both space and mechanisms to fight leakage. Both are energetically costly. Moreover, increase in size may result in a loss as the gross utilization rate increases as R^β, $2 < \beta < 3$, while the diffusive flux increases only as R, the cell radius (Jumars, 1993).

Light adaptation has it costs too. A cell has to invest in pigment assimilation in order to adapt to lower light. Thus it is understandable that the cell does not adapt to the night darkness but does adapt to the low light level of a cloudy day or a location in the deep part of the euphotic zone. If the cell does not have the nutrient to assimilate, it has no reason to adapt and thus it is sensible to assume that the adaptation is connected to the nutrient availability, as was assumed in the model.

If the above hypothesis holds, it has the potential to simplify ecological models that describe phytoplankton growth in relation to physical processes. The integrated amount of light during the day is the important parameter that will determine the cell production. Below the light threshhold production will be light limited and above it growth will be very much the same for all cells (assuming the same uptake for all cells). In the field one would have to analyze cell contents before dusk to determine their potential production.

Physics of chlorophyll distribution

In this section an equation for the distribution of chlorophyll (or any other characteristic that photoadapts, such as α in the model of the previous section) in the upper ocean is derived. It is important to derive equations connecting biomass and the more commonly measured chlorophyll in order to model upper ocean phytoplankton dynamics.

212

Derivation of the equation

Let Γ denote the amount of chlorophyll per unit weight of phytoplankton. Let the change in Γ be regulated by a first-order reaction (as (6) above):

$$\frac{\partial \Gamma}{\partial t} = \gamma \left(\Gamma_\infty(I, N) - \Gamma \right) \tag{8}$$

where Γ_∞ denotes the value towards which Γ relaxes with a time scale γ^{-1} and which will be a function of light I, (itself a function of depth and time) and nutrient availability N (as there is no point to produce more chlorophyll if there isn't a nutrient to assimilate). In this study, however, the effect of a variable nutrient field will be ignored, and light will be assumed to be the only limiting resource. Another possible formulation, ignored here, is of a logistic type, i.e. $\Gamma_t = \gamma \Gamma \frac{(\Gamma_\infty - \Gamma)}{\Gamma_\infty}$ (eq. (7) above).

Let n be the concentration of phytoplankton (in unit weight/ unit volume/unit of chlorophyll). The evolution of n is described by a conservation equation:

$$\frac{Dn}{Dt} = \mu n - s + \nabla \left(K \nabla n \right) \tag{9}$$

where s denote loss term (e.g. predation), K the diffusion coefficient and μ the growth rate, which depends on Γ, since we assume that the phytoplankton adapts in order to maximize it.

Expanding the material derivative to include the dependence of n on Γ and using (8) gives:

$$\frac{Dn}{Dt} = \frac{\partial n}{\partial t} + \nabla(\vec{u}n) + \frac{\partial}{\partial \Gamma} \left[n\gamma \left(\Gamma_\infty - \Gamma \right) \right] = \mu n - S + \nabla \left(K \nabla n \right) \tag{10}$$

where $\vec{u}$ denotes the flow field (including velocities relative to the free flow such as those induced by sinking and swimming).

The mass of phytoplankton per unit volume is given by:

$$\bar{n} \equiv \int_{\Gamma_{min}}^{\Gamma_{max}} n d\Gamma,$$

and

$$\bar{\Gamma} \equiv \frac{\int_{\Gamma_{min}}^{\Gamma_{max}} \Gamma n d\Gamma}{\bar{n}}.$$

$\bar{n}\bar{\Gamma}$ is the thus the mass of chlorophyll per unit volume. Integrating (10) with respect to Γ and using (8), we derive an equation for the rate of change of phytoplankton mass:

$$\frac{\partial \bar{n}}{\partial t} + \nabla(\vec{u}\bar{n}) = \int_{\Gamma_{min}}^{\Gamma_{max}} \mu n d\Gamma - \int_{\Gamma_{min}}^{\Gamma_{max}} s d\Gamma + \nabla(K\nabla\bar{n}) \tag{11}$$

Assuming that predation does not depend on Γ (sensible unless Γ correlates with size), one can define $s = S(\bar{n})n$ which when inserted to the RHS of (11) gives $S(\bar{n})\bar{n}$. An equation for the chlorophyll concentration ($\eta \equiv \bar{\Gamma}\bar{n}$) evolution is found by multiplying (10) with Γ and integrating with respect to Γ:

$$\frac{\partial \eta}{\partial t} + \nabla(\vec{u}\eta) - \gamma(\Gamma_{\infty}\bar{n} - \eta) = \int_{\Gamma_{min}}^{\Gamma_{max}} \Gamma \mu n d\Gamma - S(\bar{n})\eta + \nabla(K\nabla\eta) \tag{12}$$

which is coupled to (11).

It is sensible to assume that μ, the growth rate, follows a curve of the monod type:

$$\mu = \frac{\mu_{max}(z,t)\Gamma}{(K_{\mu} + \Gamma)} \tag{13}$$

which can be fitted locally fitted to straight lines, generating only one higher moment in eq. (11) (a closure problem). Although the system of equation (11)-(12) are mass balance equation of the directly measured variables (chl/volume and biomass/volume), the light -adaptation literature has focused on the changes of the $\bar{\Gamma}$ itself, the concentration of chlorophyll per unit of phytoplankton mass. An equation for $\bar{\Gamma}$ can be derived by subtracting from (12) $\bar{\Gamma}$ times (13) and dividing by $\bar{n}$, resulting in:

$$\frac{\partial \bar{\Gamma}}{\partial t} + \vec{u}\nabla\bar{\Gamma} - \gamma(\Gamma_{\infty} - \bar{\Gamma}) = \frac{1}{\bar{n}} \int_{\Gamma_{min}}^{\Gamma_{max}} (\Gamma - \bar{\Gamma})\mu n d\Gamma +$$

$$\nabla K(\nabla\bar{\Gamma}) + 2K(\nabla ln(\bar{n})) \cdot \nabla\bar{\Gamma} \tag{14}$$

notice that the growth rate term disappears if $\mu \neq \mu(\Gamma)$. To my knowledge the derivation presented here has not appeared elsewhere. The resulting equation (14), without the

growth-term contribution was given in Lande and Lewis (1989) (hereafter LL89), who have studied different growth related photoadaptive parameters other than chlorophyll.

Simplifying equation (14) to one dimension (z), assuming no time dependence, no advection, constant $\bar{n}$, constant K, and $\Gamma_\infty = a + bz$ (i.e. i.e linearly related to $\log(I)$), results in the solution (LL89):

$$\Gamma(z) = \Gamma_\infty + bLe^{-z/L}, \quad L = \sqrt{K/\gamma} \tag{15}$$

where a no-flux boundary condition at the surface was used. The effect of adaptation is seen to be confined to a top boundary layer. The effect deepens with the level of turbulence and the speed of adaptation.

Working with chlorophyll per unit mass of phytoplankton (14) rather than per unit volume of water (12) has the advantage that the signal from region of lower productivity can be extracted from a larger volume of water. Two drawbacks of this model is its neglect of the nutrient dynamics (unless conserved in μ) and the hidden assumption of no perturbation from average values measured due either to biological variability (μ, γ, s and motility) and the environment ($\vec{u}$ and K). Investigating only one genus of phytoplankton may limit the biological variability, making the above eulerian model more applicable.

In the next section some potential effects of the flow variability on phytoplankton growth-rate and photoadaptive variable distribution are presented. Two different models are presented; the first model is eulerian, based on a simplified version of the above model, and is used to solve for the distribution of a photoadaptive parameter and its dependence on background waves frequency. The second model is a lagrangian model, following the cell movement in a wavy flow, in which the effect of the waves on the mean growth rate is sought. The cells are assumed to be passive tracers of the flow.

Eulerian model with fluid flow variability

Let $\vec{u}$ be a two dimensional (x,z) flow field representing an internal wave field around a thermocline,

$$\Psi = -caze^{-\alpha z}\cos(kx - \omega t)$$

where c is the phase speed and a a nondimensional amplitude.

$$(u, w) = (-\Psi_z, \ \Psi_x) = \left[cae^{-\alpha z}(1 - \alpha z)\cos(kx - \omega t), waze^{-\alpha z}\sin(kx - \omega t) \right]$$

with the thermocline located at $z = \frac{1}{\alpha}$. Assuming that $\bar{n}$ and K are constant and that $\mu \neq \mu(\Gamma)$, (14) becomes:

$$\frac{\partial \bar{\Gamma}}{\partial t} + \vec{u}\nabla\bar{\Gamma} - \gamma\left(\Gamma_\infty - \bar{\Gamma}\right) = K\nabla^2\bar{\Gamma} \tag{16}$$

Decomposing $\bar{\Gamma}$ into its x-average $< \bar{\Gamma} >^x$ and its perturbation $\bar{\Gamma}'$, with $\bar{\Gamma}' \ll < \bar{\Gamma} >^x$.

Assuming that the velocity field is a perturbation quantity (i.e. of the same order as $\bar{\Gamma}'$) and dropping the bars from the variables, the equation for $< \Gamma >^x$ is given by the x-average of equation (16):

$$\frac{\partial < \Gamma >^x}{\partial t} + \frac{\partial < w\Gamma >^x}{\partial z} = \gamma\left(\Gamma_\infty - < \Gamma >^x\right) + K_V\frac{\partial < \Gamma >^x}{\partial z^2} \tag{17}$$

Without the second term on the LHS (assumed of smaller order) the long-time solution to (17) is given by (15).

Subtracting (17) from (16):

$$\frac{\partial \Gamma'}{\partial t} + \left\{ J(\Psi, \Gamma') - < J(\Psi, \Gamma') >^x \right\} + w\frac{\partial < \Gamma >^x}{\partial z} = \gamma\Gamma' + K_H\frac{\partial^2\Gamma'}{\partial x^2} + \left\{ K_V\frac{\partial^2\Gamma'}{\partial z^2} \right\} \tag{18}$$

t The terms in the curly brackets are neglected, assuming that they are of smaller order than this equation (which is first order in the perturbation). Further assuming that after the transient dies, $\Gamma' \propto e^{i(kx - \omega t)}$, and substituting the solution given by (15):

$$\frac{\partial \Gamma'}{\partial t} + (\gamma + K_H k^2)\Gamma' = -w\left[b(1 - e^{-z/L} \right] \tag{19}$$

Integrating in time and looking for the long time solution:

$$\Gamma' = -b(1 - e^{-z/L})\frac{waze^{-\alpha z}}{(\delta^2 + \omega^2)}\left(\delta\sin(kx - \omega t) + \omega\cos(kx - \omega t) \right)$$

216

where $\delta \equiv (K_H k^2) + \gamma$. Thus:

$$< \Gamma' w >^x = -b(1 - e^{-z/L}) \frac{\omega^2 \delta a^2 z^2 e^{-2\alpha z}}{2(\delta^2 + \omega^2)},$$

so that the second term in (17), according to this solution, is:

$$\frac{\partial < \Gamma' w >^x}{\partial z} = \frac{b\omega^2 \delta a^2}{2(\delta^2 + \omega^2)} e^{-2\alpha z} z \left(2(\alpha z - 1)(1 - e^{-z/L}) - \frac{z}{L} e^{-z/L} \right)$$

This term (divided by γ) gives the correction to Γ below the top boundary layer (assuming the thermocline is deeper). It scales like the perturbation amplitude squared, and its neglect is thus consistent with the solution (15). The correction term is maximal for high frequency waves (as the cells do not then have time to adapt) and decreases with K_H (since horizontal gradients are smoother). Although an exact solution of (17) was not derived the range of validity of the solution of (17) with a neglected term and its estimated error was found. It is thus expected that the adaptation will be surface trapped depending more on turbulent mixing (K) than internal waves amplitudes (a).

Adaptation and internal waves, a lagrangian view:

In this section an effect of fluid motion and light adaptation on the phytoplankton growth-rate is investigated. A simple example illustrating it is as follows: suppose that a parcel of water is oscillating in a regular manner (e.g. due to the above internal waves). Assuming that the photoadaptive property is proportional to $\log(I)$ (LL89):

$$\Gamma_\infty = a + b(z - z_0). \tag{20}$$

Solving (8):

$$\Gamma = \gamma e^{-\gamma t} \int_0^t \Gamma_\infty e^{\gamma t'} dt' + \Gamma(t = 0)e^{-\gamma t}. \tag{21}$$

Assuming the fluid parcel oscillate aound z_0, $z = z_0 + B\sin(\omega t)$, and defining $C = bB$, (21) gives:

$$\Gamma = \gamma e^{-\gamma t} \int_0^t \left(a + C\sin(\omega t) \right) e^{\gamma t'} dt' + \Gamma_0 e^{-\gamma t} =$$

217

$$a(1 - e^{-\gamma t}) + \frac{\gamma C}{(\gamma^2 + \omega^2)} \left(\gamma \sin(\omega t) - \omega \cos(\omega t) - \gamma e^{-\gamma t} \right) + \Gamma_0 e^{-\gamma t},$$

which becomes when $t \to \infty$ (with e-folding scale γ):

$$\Gamma = a + \frac{\gamma C}{(\gamma^2 + \omega^2)} \left(\gamma \sin(\omega t) - \omega \cos(\omega t) \right) = < \Gamma > + \Gamma'. \tag{22}$$

The mean of the photosynthetic trait is equal to its value at its mean position, a result which is not surprising due to the linearity of Γ_∞ with z. Still the effect over the growth rate of the deviation from the mean can be appreciable. As an example, assume that the photosynthetic trait Γ is the α of the growth rate formula (Platt et. al., 1980):

$$P = p_m \left(1 - e^{-\frac{\alpha}{p_m} I} \right); \quad I = I_0 e^{-kz},$$

for which both a and b (in (21)) are positive (LL89). In order to make it mathematically tractable, it is assumed that the parcel is deep enough (below 30m) so that:

$$P \approx \alpha I_0 e^{-kz}.$$

Calculating the average growth rate:

$$< P > = I_0 e^{-kz_0} < \alpha e^{kB \cos(\omega t)} > = I_0 \frac{e^{-kz_0}}{T} \int_0^T < \alpha > e^{-kB \cos(\omega t)} + \alpha' e^{-kB \cos(\omega t)} dt.$$

Using (17) gives:

$$< P > = I_0 e^{-kz_0} \left[a \mathcal{I}_0(kB) + \frac{\gamma \omega C}{(\gamma^2 + \omega^2)} \mathcal{I}_1(kB) \right],$$

where $\mathcal{I}_n$ is the modified Bessel function of the n-th order. The first term which was analyzed by Flierl (personal communication) is a production increase (relative to the non-wavy case) due to the nonlinear light structure. The second term, due to the light adaptation, is also a production increase ($b > 0$), largest when $\omega = \gamma$, the natural adaptation time of the system. It is natural to expect that the photoadaptive response time will be of the same oscillation frequency as the vertical motion experienced by the phytoplankton since doing

it will maximize their growth. This strategy will be useful for coherent motions, such as seiches and internal tides (with time scales of 6-24 hours, similar to observed time scale for light adaptation).

Acknowledgements

Bibliography

Cullen, J. J. and M. R. Lewis, 1988: The kinetics of algal photoadaptation in the context of vertical mixing. J. Plank. Res., 10, 1039-1063.

Denman K. L. and J. Mara, 1986: Modeling the time dependent adaptation of phytoplankton to fluctuating light. In J. C. J. Nihoul (ed.) Marine interfaces ecohydrodynamics. Elsevier, Amsterdam, p. 341-359.

Harding, L. W., Meeson, B. W., Prezelin, B. B. and B. M. Sweeney, 1981: Diel periodicity of photosynthesis in marine phytoplankton. Mar. Biol., 61, 95-105. Jumars P., 1993:

Jumars, P. A., 1993: Concepts in biological oceanography. Oxford U. Press, New-York, 348pp.

Lande, R. and M. R. Lewis, 1989: Models of photoadaptation and photosynthesis by algal cells in a turbulent mixed-layer. Deep-sea Res., 36, 1161-1175.

Mara, J., 1978a: Effect of short-term variations in light intensity on photosynthesis of a marine phytoplankter: a laboratory simulation study. Mar. Biol., 46, 191-202.

Mara, J., 1978b: Phytoplankton photosynthetic response to vertical movement in a mixed layer. Mar. Biol., 46, 203-208.

Parsons, T. R., Takahashi, M. and B. Hargrave, 1984: Biological oceanographic processes, 3rd edition. Pergamon Press, Oxford, 330pp.

Platt, T., Gallegos, C. L. and W. G. Harrison, 1980: Photoinhibition of photosynthesis in natural assemblages of marine phytoplankton. J. Mar. Res., 38, 687-701.

Prezelin B. B., 1992: Diel periodicity in phytoplankton productivity. Hydrobiologia, 238, 1-35.
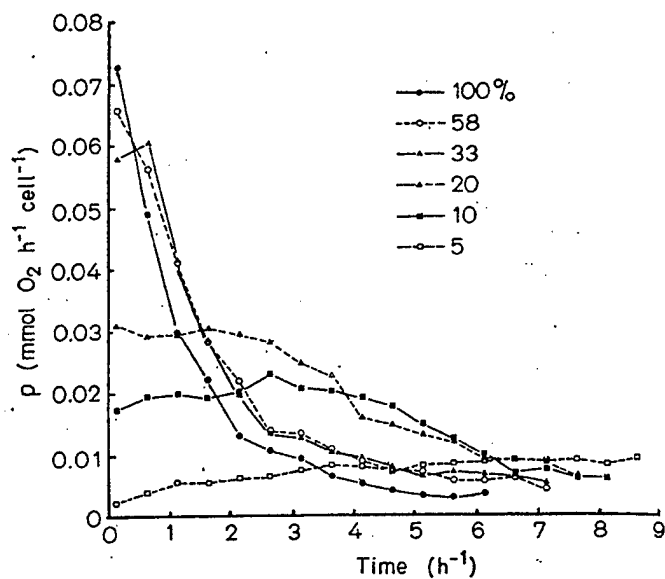
Fig. 1. O_2 production (p) over time for various
light levels. After 12 hr of dark light was
turned on at time t=0. 100%=1500μE/m²/sec.
From Mara (1978b).
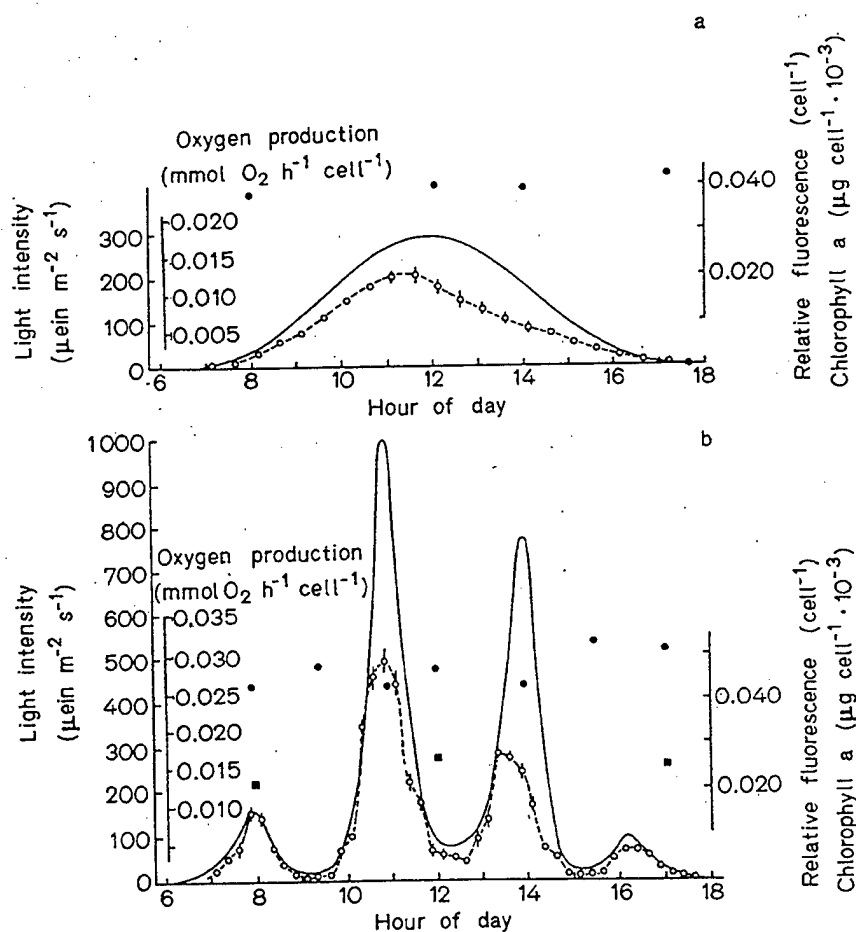
Fig. 2. O_2 production over time for both diurnally varying and fluctuating
regime. The total light intensity is the same in both experiments and so is
the time-integrated production. From Mara (1978a).
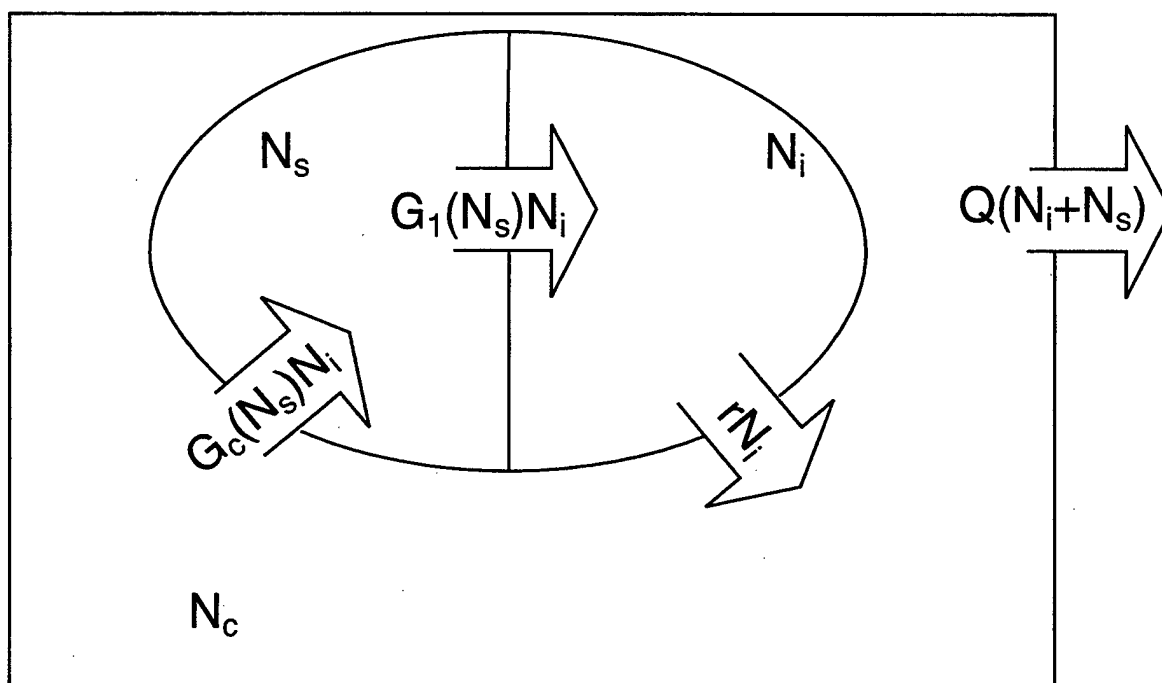
Fig. 3. Schematics of the experiment model. Arrows denote fluxes, N, the amount of nutrient in each compartment, G, the uptake and Q the harvest rate.
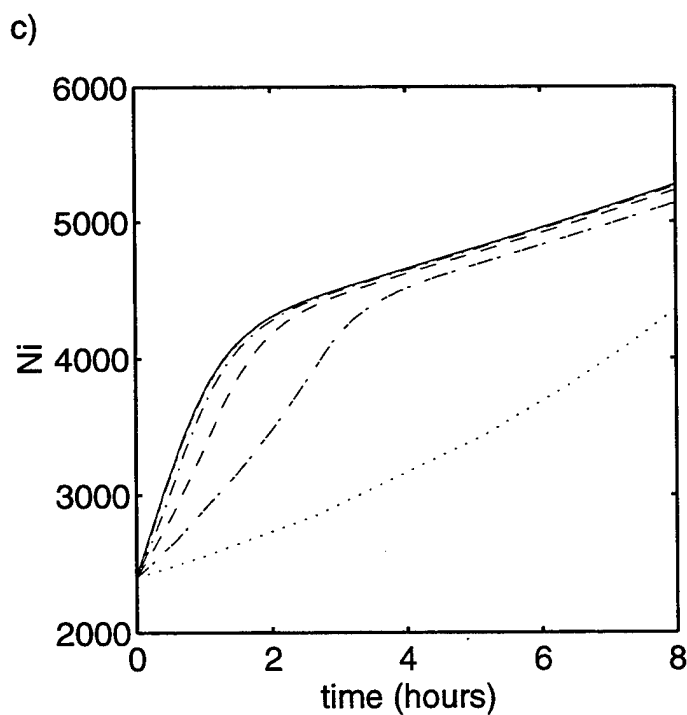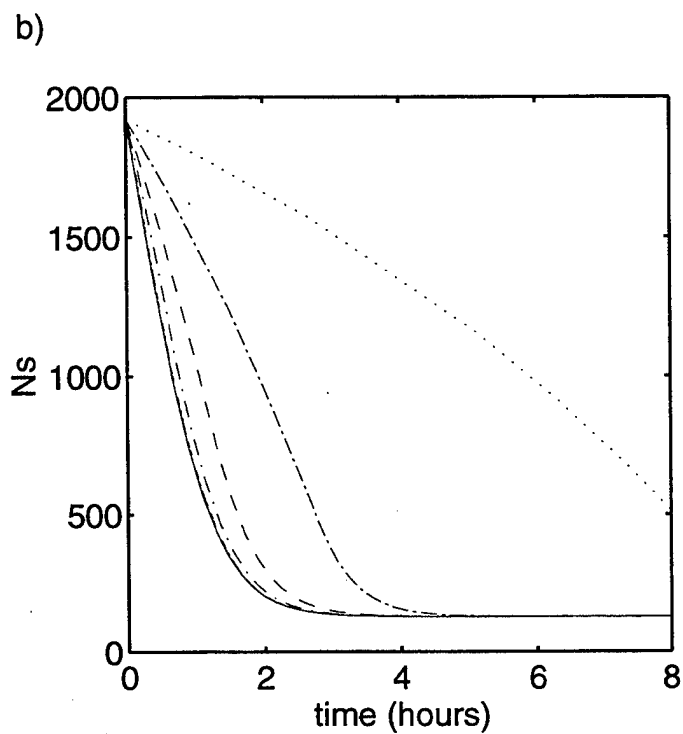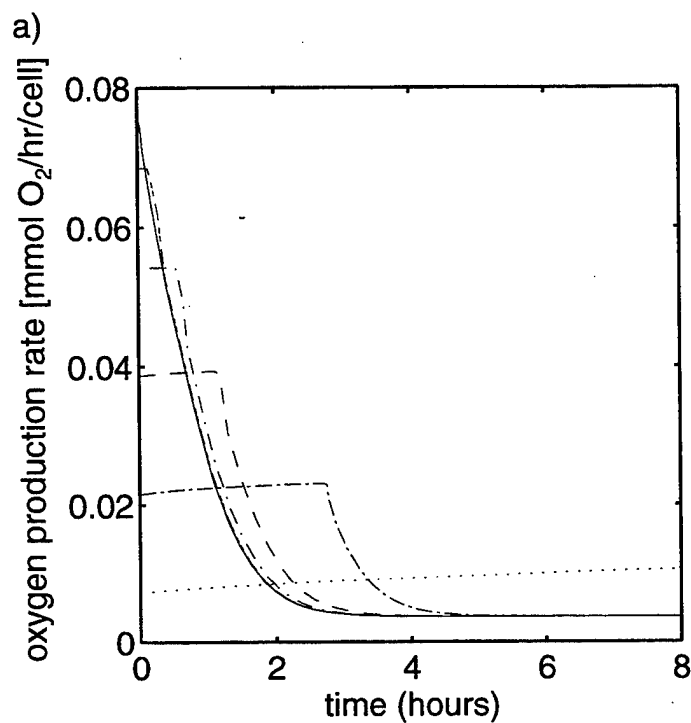
Fig. 4. Results of numerical model. a) O_2 production rate [mmol/hr/cell] (to be compared with Fig. 1). b)Stored nutrient [mmol O_2]. c) Incorporated nutrient [mmol O_2]. For light intensities: 100% (solid), 58% (dash dot dot), 33% (dash dot), 20% (dash), 10%(dash dot) and 5% (dot) of 1500µE/m²/sec.

Allele Frequency Clines
in a Population Genetics Model

Amanda Cohen
Univ. of Washington

INTRODUCTION

Spatial gradients or clines in allele frequencies are important for the maintenance of genetic diversity in natural populations. These clines can develop in diallelic populations living in spatially heterogeneous environments, if the two alleles have different selection coefficients in different regions of the domain, and the migration rate is low relative to the rate of selection.

Several authors have modeled the formation and maintenance of allele frequency clines. Slatkin (1973), for example, modeled a large, continuously distributed, randomly mating population subject to natural selection. He derived a characteristic length scale for his model, $l_c = l/\sqrt{s}$ (where s is the strength of the selection), and suggested that significant gene frequency variations cannot develop at scales less than this characteristic length. Similarly, Nagylaki (1975) defined a dimensionless parameter, $k^2 = 2sa^2/\sigma^2$ (where s is again the strength of selection, a is the length scale of habitat variation, and σ is a measure the variance in migration). He then argued that, subject to certain simplifying assumptions, a cline can only exist if $k > \tan^{-1}\alpha$, where α is the ratio of the selection coefficients on either side of the (one-dimensional) domain. Keller (1984) studied cline formation in a one, two or three dimensional region. He included the effects of diffusion and selection, and he developed an asymptotic solution describing the steady state spatial structure of the cline, for the case in which selection is much stronger than diffusion. Keller (1984) also provides a brief review of several other cline formation models, including the two-dimensional version of Nagylaki's (1975) formulation.

The model developed here follows a population of diploid, randomly mating organisms in a one-dimensional, bounded domain, to investigate the effects of diffusion and spatially varying selection on allele frequencies at a single, diallelic locus, neglecting the effects of mutation and genetic drift. Unlike many earlier models of cline formation, this model considers the diffusion of individuals rather than alleles, so the population is allowed to stray from Hardy-Weinberg equilibrium.

The model consists of six equations – a continuous reaction-diffusion equa-

tion and a discrete reproduction equation for each of three genotypes. The reproduction equations are applied at the end of each generation; the continuous equations are applied between reproductive events. Eigenmode solutions are found for the three continuous equations, and the first (or first few) modes from each solution are substituted into the reproduction equations to generate a discrete map. The map describes the number of individuals of each genotype at the start of a generation as functions of the population size and genotype distribution at the end of the previous generation. The behavior of the map is evaluated for various forms of the selection functions. With spatially homogeneous selection, genetic diversity can only be maintained if the heterozygote has a selective advantage. Results with spatially varying selection depend on the relative length of the generation time, the shape of the selection functions, and the spatial symmetry of the system of equations.

FORMULATION

The model considers a diallelic population with discrete, non-overlapping generations distributed over a bounded, one-dimensional domain. The number of individuals of each genotype evolves continuously between reproductive events according to the following modified reaction-diffusion equation:

$$\frac{\partial \rho_i(x,t)}{\partial t} = -\mu_i(x)\rho_i(x,t) + \kappa \frac{\partial^2 \rho_i(x,t)}{\partial x^2}, \qquad i = 1,2,3 \qquad (1)$$

with no-flux conditions applied at the boundaries of the domain:

$$\frac{\partial \rho_i(x,t)}{\partial x} = 0. \qquad (2)$$

In equation (1), $\rho_i(x,t)$ is the number of individuals of genotype i, $\mu_i(x)$ (hereafter referred to as the selection function) describes the spatially varying death rate of genotype i, and $\kappa(\partial^2 \rho_i/\partial x^2)$ accounts for the diffusion of individuals.

Note that the diffusion rate, κ, is spatially homogeneous and independent of genotype. The selection function varies spatially and depends on genotype, but does not include density dependence – the death rate of a given genotype is independent of the total population size.

At the end of each generation, the organisms reproduce. The number of

individuals of each genotype in the new generation is given by:

$$\rho_1'(x,0) = \frac{(f_1\rho_1(x,T) + (f_2/2)\rho_2(x,T))^2}{N(x,T)}\beta\exp\left(-\frac{\alpha}{\beta}N(x,T)\right) \quad (3)$$

$$\rho_2' = \frac{(2f_1f_3\rho_1\rho_3 + f_1f_2\rho_1\rho_2 + f_2f_3\rho_2\rho_3 + (f_2^2/2)\rho_2^2)}{N}\beta\exp\left(-\frac{\alpha}{\beta}N\right) \quad (4)$$

$$\rho_3' = \frac{(f_3\rho_3 + (f_2/2)\rho_2)^2}{N}\beta\exp\left(-\frac{\alpha}{\beta}N\right) \quad (5)$$

where $\rho_i'(x,0)$ refers to the number of individuals of genotype i at the start of the next generation, and $\rho_i(x,T)$ is the number of individuals of genotype i at the end of the current generation. In these equations, ρ_1 and ρ_3 denote homozygous individuals, and ρ_2 denotes heterozygous individuals. f_1, f_2 and f_3 are parameters describing the relative fecundity of each genotype; T is the length of a generation; $N(x,T)$ is the total population at the end of a generation $(N(x,T) = \rho_1(x,T) + \rho_2(x,T) + \rho_3(x,T))$; and α and β are parameters regulating the growth rate of the total population. The arguments of the functions in equations (4) and (5) are left out to aid in reading the equations; the arguments are the same as in equation (3).

A brief explanation of the rationale behind equations (1) - (5) is warranted. Equation (1) is similar to the equations used by Nagylaki (1975) in his analysis of cline formation in the presence of selection, advective drift and diffusion. In the present analysis, only selection and diffusion are considered. Equations (3) - (5) are modified from Webb (1981). If f_1, f_2 and $f_3 = 1$, these equations give the familiar Hardy-Weinberg ratios for the genotypes at the start of the next generation. Inclusion of the parameters f_1, f_2 and f_3 allows for more complicated dynamics, because differential fecundity and assortative mating can be approximated. A more accurate representation of differential fecundity and assortative mating would require six parameters – one for each possible type of mating pair. Only three parameters are used here, to simplify the analysis.

The exponential terms in equations (3) - (5), modified from the Ricker equation for population growth (Murray, 1993), reduce the rate of population growth for large N. These terms affect the population dynamics but do not alter the genetics, because the three genotypes are affected identically. The use of density-dependent exponential terms in the reproduction equations, (3) - (5), and not in the reaction-diffusion equation, (1), reflects the assumption that density effects (such as nutrient limitation) impact individuals far

more at the instant of reproduction than during non-reproductive periods.

SOLUTION METHOD

The solutions to the selection-diffusion equation (1) can be expressed as infinite sums of eigenmodes:

$$\rho_i(x,t) = \sum_j a_{i,j} \exp\left(-\lambda_{i,j} t\right) \psi_{i,j}(x), \qquad i = 1,2,3. \qquad (6)$$

Plugging equation (6) into equations (1) and (2), and scaling out the diffusivity, κ, (i.e. redefining $\mu_i(x)$ and $\lambda_{i,j}$), one finds that the eigenvalues, $\lambda_{i,j}$, and eigenfunctions, $\psi_{i,j}$, must satisfy

$$\frac{\partial^2 \psi_{i,j}}{\partial x^2} = \left(\mu_i(x) - \lambda_{i,j}\right) \psi_{i,j}, \qquad i = 1,2,3 \qquad (7)$$

subject to the no-flux boundary conditions

$$\frac{\partial \psi_{i,j}}{\partial x} = 0, \qquad i = 1,2,3. \qquad (8)$$

From Sturm-Liouville theory, it can be shown that for the partial differential equation (1), if the selection function $\mu_i(x)$ is real and positive, all the eigenvalues in equation (6) are also real and positive, and the lowest order eigenfunction, $\psi_{i,0}$, is everywhere positive.

Because the eigenvalues in equation (6) are ordered along the positive, real axis, the higher order modes decay more rapidly than the first mode. Thus, for a sufficiently long generation time, T, the higher order modes are negligible at the end of each generation. That is, the population size and the spatial distribution of the genotypes at the end of a generation are dictated almost entirely by the amplitude and shape of the first (or first few) modes. To a first approximation, then, equation (6) can be truncated to

$$\rho_i(x,t) = a_{i,0} \exp\left(-\lambda_{i,0} t\right) \psi_{i,0}(x). \qquad (9)$$

The accuracy of this approximation is discussed below.

Plugging the simplified solution for ρ_1 at $t = T$ (equation (9) with $i = 1$) into reproduction equation (3), one finds that

$$
\begin{aligned}
\rho_1'(x,0) &= a_{1,0}' \psi_{1,0} \qquad\qquad\qquad\qquad\qquad\qquad (10) \\
&= \frac{\left(f_1 a_{1,0} \exp(-\lambda_{1,0}T)\psi_{1,0} + (f_2/2)a_{2,0}\exp(-\lambda_{2,0}T)\psi_{2,0}\right)^2}{\sum_i a_{i,0} \exp\left(-\lambda_{i,0}T\right) \psi_{i,0}} C
\end{aligned}
$$

with

$$C(x) = \beta \exp\left(-\frac{\alpha}{\beta} \sum_i a_{i,0} \exp\left(-\lambda_{i,0} T\right) \psi_{i,0}\right).$$

The equations for ρ_2' and ρ_3' have a similar form that can easily be derived from equations (4), (5) and (9).

Taking advantage of the orthonormality of the eigenfunctions, $\psi_{i,j}$, one can multiply both sides of the simplified equations for the ρ_i's by $\psi_{i,0}$ and integrate over x to derive equations for $a_{1,0}'$, $a_{2,0}'$ and $a_{3,0}'$:

$$a_{1,0}' = \frac{1}{L} \int_L \frac{C_{1_1}^2}{\sum_i a_{i,0} \exp\left(-\lambda_{i,0} T\right) \psi_{i,0}(x)} C_2 \psi_{1,0}(x) dx \qquad (11)$$

$$a_{2,0}' = \frac{1}{L} \int_L \frac{C_3}{\sum_i a_{i,0} \exp\left(-\lambda_{i,0} T\right) \psi_{i,0}(x)} C_2 \psi_{2,0}(x) dx \qquad (12)$$

and

$$a_{3,0}' = \frac{1}{L} \int_L \frac{C_{1_3}^2}{\sum_i a_{i,0} \exp\left(-\lambda_{i,0} T\right) \psi_{i,0}(x)} C_2 \psi_{3,0}(x) dx \qquad (13)$$

with $C_{1_i}(x)$, $C_2(x)$ and $C_3(x)$ given by

$$
\begin{aligned}
C_{1_i}(x) &= f_i a_{i,0} \exp\left(-\lambda_{i,0} T\right) \psi_{i,0}(x) + (f_2/2) a_{2,0} \exp\left(-\lambda_{2,0} T\right) \psi_{2,0}(x), \\
C_2(x) &= \beta \exp\left(-\frac{\alpha}{\beta} \sum_i a_{i,0} \exp\left(-\lambda_{i,0} T\right) \psi_{i,0}\right), \qquad \text{and} \\
C_3(x) &= 2 f_1 f_3 a_{1,0} a_{3,0} \exp\left(-(\lambda_{1,0} + \lambda_{3,0}) T\right) \psi_{1,0}(x) \psi_{3,0}(x) \\
&\quad + f_1 f_2 a_{1,0} a_{2,0} \exp\left(-(\lambda_{1,0} + \lambda_{2,0}) T\right) \psi_{1,0}(x) \psi_{2,0}(x) \\
&\quad + f_2 f_3 a_{2,0} a_{3,0} \exp\left(-(\lambda_{2,0} + \lambda_{3,0}) T\right) \psi_{2,0}(x) \psi_{3,0}(x) \\
&\quad + (f_2^2/2) a_{2,0}^2 \exp\left(-2\lambda_{2,0} T\right) \psi_{2,0}(x)^2.
\end{aligned}
$$

Equations (11), (12) and (13) form a 3x3 discrete map that gives, for each genotype, the amplitude of the first mode at the start of the next generation in terms of the amplitudes at the start of the current generation. This map can be iterated from one generation to the next to determine, to first order, the temporal evolution of the populations of each genotype.

The allele frequencies can be calculated from the total population size and the genotype distribution. Letting A_1 and A_2 denote the two allele types, and p and q the respective allele frequencies, and assigning ρ_1, ρ_2 and ρ_3 individuals the genotypes $A_1 A_1$, $A_1 A_2$ and $A_2 A_2$ respectively, one finds that

p and q are given by:

$$p(x,t) = \frac{\rho_1(x,t) + (\rho_2(x,t)/2)}{\rho_1(x,t) + \rho_2(x,t) + \rho_3(x,t)} \tag{14}$$

and

$$q(x,t) = 1 - p(x,t) = \frac{\rho_3(x,t) + (\rho_2(x,t)/2)}{\rho_1(x,t) + \rho_2(x,t) + \rho_3(x,t)} \tag{15}$$

Because of the modal truncations in the ρ_is, the formulae for p and q are most accurate for $t = T$ – i.e. at the end of a generation, when the contributions from the higher order modes are smallest.

To improve the accuracy of this model, one can use additional modes in equations (9). Using, for example, the first three modes in the solutions for the ρ_is results in a 9x9 discrete map similar in form to equations (11) - (13).

RESULTS

Model results are discussed for both spatially homogeneous and spatially inhomogeneous selection functions. For model runs with spatially homogeneous μ_is, the eigenfunctions and eigenvalues are found analytically. For spatially varying μ_is, the eigenfunctions are determined numerically. All map iterations are performed numerically.

In the spatially homogeneous case, only a single-mode truncation is discussed. In the spatially inhomogeneous case, results from a one-mode truncation are compared with those from a three-mode truncation, for both a long and a short generation time, T, to explore the effects of reducing the generation time.

I. Spatially Homogeneous Selection

In the first set of model runs, spatially homogeneous selection functions were used – that is, $\mu_i(x) = \mu_i$. Only single-mode truncations were investigated. With spatially homogeneous μ_is, the eigenfunction equation, (7), has sinusoidal solutions of the form

$$\psi_{i,j} = \cos(\sqrt{\lambda_{i,j} - \mu_i} x + \phi).$$

The eigenvalues, $\lambda_{i,j}$, are then defined by

$$\lambda_{i,j} = \frac{\pi^2 j^2}{L^2} + \mu_i \qquad (16)$$

in ¡which L is the length of the domain. Truncating this solution to the zeroth mode ($j = 0$; $\lambda_{i,0} = \mu_i$) results in simple, spatially homogeneous solutions for the ρ_is:

$$\rho_i = a_{i,0} \exp(-\mu_i t).$$

For this form of the ρ_is, equations (11) - (13) simplify somewhat to:

$$a'_{1,0} = \frac{(f_1 a_{1,0} \exp(-\mu_1 T) + (f_2/2) a_{2,0} \exp(-\mu_2 T))^2}{D_1} \beta \exp\left(-\frac{\alpha}{\beta} D_1\right) \quad (17)$$

$$a'_{2,0} = \frac{D_2}{D_1} \beta \exp\left(-\frac{\alpha}{\beta} D_1\right) \qquad (18)$$

and

$$a'_{3,0} = \frac{(f_3 a_{3,0} \exp(-\mu_3 T) + (f_2/2) a_{2,0} \exp(-\mu_2 T))^2}{D_1} \beta \exp\left(-\frac{\alpha}{\beta} D_1\right) \quad (19)$$

in which

$$
\begin{aligned}
D_1(x) =\ & a_{1,0} \exp(-\mu_1 T) + a_{2,0} \exp(-\mu_2 T) + a_{3,0} \exp(-\mu_3 T), \quad \text{and} \\
D_2(x) =\ & 2 f_1 f_3 a_{1,0} a_{3,0} \exp(-(\mu_1 + \mu_3)T) \\
& + f_1 f_2 a_{1,0} a_{2,0} \exp(-(\mu_1 + \mu_2)T) \\
& + f_2 f_3 a_{2,0} a_{3,0} \exp(-(\mu_2 + \mu_3)T) + (f_2/2) a_{2,0}^2 \exp(-2\mu_2 T).
\end{aligned}
$$

Interestingly, this apparently three-dimensional map is actually composed of a two-dimensional map that defines the ratios among the $a_{i,0}$s, and an independent, one-dimensional map that defines their magnitudes. The two-dimensional map rapidly approaches a steady-state solution, but the magnitudes of the $a_{i,0}$s continue to vary according to the one-dimensional map. The separation into a one- and a two-dimensional map is best illustrated by solving for the fixed points of the two-dimensional map. Taking $a_{2,0}$ as the independent variable, one can look for constants k_1 and k_3 such that $a_{1,0} = k_1 a_{2,0}$ and $a_{3,0} = k_3 a_{2,0}$. Plugging these formulae into equations (17) - (19) generates recursive equations for k_1 and k_3:

$$k'_1 = \frac{a'_{1,0}}{a'_{2,0}} = \frac{(k_1 f_1 \exp(-\mu_1 T) + (f_2/2) \exp(-\mu_2 T))}{2(k_3 f_3 + (f_2/2))} \qquad (20)$$

230

and

$$k_3' = \frac{a_{3,0}'}{a_{2,0}'} = \frac{(k_3 f_3 \exp(-\mu_3 T) + (f_2/2)\exp(-\mu_2 T))}{2(k_1 f_1 + (f_2/2))}. \qquad (21)$$

From these equations, it is evident that $k_1 k_3 = \frac{1}{4}$ for every generation after the first. Substituting this relationship into equation (20) one finds that

$$k_1' = \frac{2k_1^2 f_1 \exp(-\mu_1 T) + k_1 f_2 \exp(-\mu_2 T)}{2k_1 f_2 \exp(-\mu_2 T) + f_3 \exp(-\mu_3 T)} \qquad (22)$$

The fixed points of this equation can be found by setting the left-hand-side equal to k_1. If f_1 and f_3 are not equal, these fixed points have the form

$$k_1 = \frac{F_3 - F_2}{2(F_1 - F_2)}$$

and

$$k_3 = \frac{F_1 - F_2}{2(F_3 - F_2)},$$

where $F_j = f_j \exp(-\mu_j T)$. The nature of the solutions depends on the ratios $\frac{f_1 \exp(-\mu_1 T)}{f_2 \exp(-\mu_2 T)}$ and $\frac{f_2 \exp(-\mu_2 T)}{f_3 \exp(-\mu_3 T)}$. Diagrams of the evolution of k_1 over time are shown in figure 1. In cases in which k_1 goes to 0 or ∞, the heterozygote and one of the two homozygotes (ρ_1 or ρ_3) goes extinct. A stable fixed point, in which both ρ_1 and ρ_3 survive, only exists if

$$\frac{f_1 \exp(-\mu_1 T)}{f_2 \exp(-\mu_2 T)} < 1 < \frac{f_2 \exp(-\mu_2 T)}{f_3 \exp(-\mu_3 T)}.$$

This result, which has been discussed by many authors (see, for example, Hartl and Clark, 1989) indicates that genetic diversity present in the system can only be maintained if the fitness of the heterozygote is greater than the fitness of either homozygote – a situation called overdominance.

Although the ratios among the $a_{i,0}$s approach steady-state after only a few generations, the magnitudes of the $a_{i,0}$s continue to vary according to the following one-dimensional map:

$$a_{2,0}' = \frac{D_3 a_{2,0}}{k_1 \exp(-\mu_1 T) + \exp(-\mu_2 T) + k_3 \exp(-\mu_3 T)} D_4$$

in which

$$
\begin{aligned}
D_3 &= 2k_1 k_3 f_1 f_3 \exp((\lambda_1 + \lambda_3)T) + k_1 f_1 f_2 \exp((\lambda_1 + \lambda_2)T) \\
&\quad + k_3 f_2 f_3 \exp((\lambda_2 + \lambda_3)T) + (f_2^2/2)\exp((\lambda_1 + \lambda_3)T), \qquad \text{and} \\
D_4 &= \beta \exp\left(-\frac{\alpha}{\beta}(k_1 \exp(\lambda_1 T) + \exp(\lambda_2 T) + k_3 \exp(\lambda_3 T)) a_{2,0}\right)
\end{aligned}
$$

This equation describes a positive, unimodal map (see, e.g., Rasband, 1990), that leads to chaotic population dynamics for certain values of β, but the interesting genetics are already resolved. The genotype frequencies are fixed in the ratios k_1:1:k_3 everywhere in the domain. The behavior of the map in this case is similar to the case discussed below, of a one-mode truncation with symmetric, spatially varying selection functions.

II. Spatially Inhomogeneous Selection

Two cases of inhomogeneous selection functions were tested: a symmetric case, with $\mu_1(x) = (x+1)/2$, $\mu_2(x) = 1/2$ and $\mu_3(x) = (-x+1)/2$, and an asymmetric case, with $\mu_1 = \mu_2 = (x+1)/2$ and $\mu_3 = (-x+1)/2$. Symmetric selection functions represent a situation in which there is no dominance – the fitness of the heterozygote is the average of the homozygotes' fitnesses. These asymmetric selection functions represent total dominance – the fitness of the heterozygote is equal to the fitness of the A_1A_1 homozygote.

For these model runs, the domain extended from $x = -1$ to $x = 1$. The first three eigenfunctions for each case were calculated numerically, using equation (7). Both the selection functions and the eigenfunctions for the symmetric case are shown in figure 2. Those for the asymmetric case are shown in figure 3.

With inhomogeneous selection functions, the temporal evolution of the spatial structure of each genotype depends strongly on the generation time, T. In the following discussion, the mathematical importance of the generation time is first addressed. Model results for a single-mode truncation with a long generation time are then described. Finally, the effects of a short generation time are explored in a comparison of results with one- and three-mode truncations. All the results are discussed for both symmetric and asymmetric selection functions, with f_1, f_2 and f_3 set to 1. Note that the qualitative behavior of the model with asymmetric fecundity parameters, f_1, f_2 and f_3, would be similar to that with asymmetric μ_is, since the f_js and λ_js occur together in equation (10).

The Importance of the Generation Time

With inhomogeneous selection, the behavior of this model depends strongly on the generation time, T. At each birth event, all of the eigenmodes are present. For a population with a long generation time, the higher eigenmodes decay away before the next birth event, so the population structure

at the end of a generation is determined entirely by the shapes and amplitudes of the first eigenfunctions. As a result, the model behaves almost identically with either a single- or a multiple-mode truncation.

With shorter generation times, however, the higher order modes do not decay completely between generations. The implications for the genetic dynamics of this additional mathematical complexity can be investigated by comparing results with single- and multiple-mode truncations. Results of such a comparison are discussed below.

For either a long or a short generation time, the error introduced by a single-mode truncation can be estimated by investigating the ratio of the amplitudes of the first and second eigenmodes at the end of a generation: $\exp\left(-(\lambda_{i,2} - \lambda_{i,1})T\right)$. For the long generation time results discussed below, the average value of this ratio is 0.0064, indicating that a single-mode truncation is a reasonable approximation in this case.

Long T, Symmetric Selection, One-mode Truncation

With symmetric μ_is, the apparently three-dimensional map that results from a single-mode truncation (equations (11) - (13)) is in fact only two-dimensional, which simplifies the genetic dynamics. A bifurcation diagram of $a_{1,0}$ versus β shows a period doubling cascade, indicating that the population dynamics are chaotic for certain ranges of β (figure 4). Even for values of β for which the population dynamics are chaotic, however, the dynamics of the genetics are quite simple. For example, as shown in figure 5, $a_{1,0}$ iterates chaotically for $\beta = 2.5$, so the total population size varies considerably from generation to generation. In contrast, the map in figure 6 indicates that $a_{3,0}$ is locked to $a_{1,0}$, even as $a_{1,0}$ iterates chaotically. Because $a_{1,0}$ and $a_{3,0}$ are locked to each other, temporal change in allele frequencies results from the relative stretching of $\psi_{1,0}$ and $\psi_{3,0}$ on opposite sides of the domain, as $a_{1,0}$ and $a_{3,0}$ vary. An average of the frequency of allele A_1 over 100 generations is shown in figure 7. The spatial structure of this average is fixed by the shapes of $\psi_{1,0}$ and $\psi_{3,0}$, independent of the value of β.

In this simplest inhomogeneous case, the variance in the allele frequencies is zero at the center of the domain, because the first eigenfunctions cross and $f_1 = f_3$, so the allele frequencies are always 0.5 (figure 8). At the edges of the domain, the variance increases with increasing β, reflecting the fact that chaotic population dynamics increase the variance in allele frequencies at a point in space, even if the time-averaged allele frequencies are fixed.

Long T, Asymmetric Selection, One-mode Truncation

With asymmetric μ_is, the map for the a_is produced by a single mode truncation is truly three-dimensional (figures 9 and 10). As a result, the dynamics of the genetics are more complicated than in the symmetric case, although the *shape* of the cline in allele frequencies is still fixed by the shape of the first eigenfunctions. As might be expected, the average frequency of allele A_1 is lower everywhere in the domain (for all values of β) in the asymmetric than the symmetric case. This result can be understood if one recalls that the asymmetric case represents total dominance of the A_1 allele – essentially, allele A_2 has a greater fitness on the side of the domain in which A_1 is favored, because A_2 can "hide" in heterozygous individuals. With increasing β, the average frequency of allele A_1 decreases (figure 11), suggesting that chaotic population dynamics benefit the recessive allele, even in regions of the domain in which the dominant allele is favored.

In this asymmetric case, the variance of the allele frequencies increases with increasing β. This increase is most pronounced on the right side of the domain, where the recessive allele is favored (figure 12).

Short T, Symmetric Selection
Comparison of One- and Three-mode Truncations

With a long generation time, results with one- or three-mode truncations are identical. With a short generation time, however, a three-mode truncation allows for more complicated spatial structure of the allele frequencies, because the higher order modes do not decay completely before the end of each generation. With symmetric selection functions, the genetic dynamics for a three-mode truncation depend on the value of β. For low values of β ($\beta <\sim 2.35$), the behavior of the model is similar to that for a one-mode truncation. For higher values of β, though, $a_{3,0}$ is no longer locked to $a_{1,0}$, and the shape of the cline in allele frequencies varies temporally. This change in the dynamics of $a_{3,0}$ and $a_{1,0}$ is obvious in the bifurcation diagram shown in figure 13. The time-averaged shape of the resulting cline in allele frequencies is shown in figure 14, for several values of β. Interestingly, as the value of β increases, the slope of the cline decreases, suggesting that for a population with a short generation time, chaotic population dynamics can erase (or mask) allele frequency clines.

The variance in frequency of allele A_1 is shown in figure 15 for several values of β. As for the simpler cases with longer generation times or single-mode truncations, the variance initially increases with increasing β. To understand

234

the peak in variance obvious at $x \simeq 0.5$ for $\beta = 3.5$, one must examine more closely the inter-generational behavior of the system for that value of β. Figure 16 shows the period-two oscillations that occur in ρ_1, ρ_2 and ρ_3 for $\beta = 3.5$. The reaction-diffusion dynamics are evident: when the population is large on the left side of the domain, population growth is slowed in that area, and diffusion toward the right is maximized. As a result, in the next generation, the population is large on the right side of the domain, population growth is slowed there, and diffusion toward the left is maximized. These dynamics are essentially identical for all three genotypes, in spite of the differences in the spatial structure of the selection functions. In effect, the time scales of diffusion and total population growth are sufficiently short that selection plays little or no role in the genetic dynamics; the system is physically dominated.

The odd peak in variance at $x \simeq 0.5$ results from this reaction-diffusion oscillation in the spatial structure of the genotypes. The first and third modes of each genotype oscillate out of phase with each other. The third mode has a peak on the right side of the domain (visible in figure 2) that is obscured when the total population is large in that area but is evident (in alternating generations) when the population on the right side of the domain is nearly zero.

Short T, Asymmetric Selection
Comparison of One- and Three-mode Truncations

For a short generation time, the results of a three-mode truncation with asymmetric selection are extremely similar to those with symmetric selection. For low values of β, the time-averaged cline in gene frequency has the same spatial structure as the first eigenfunction. For higher values of β, the average cline flattens and changes shape, as spatial gradients in genotype frequency are damped, and the higher order modes begin to affect the genetic dynamics. As with symmetric selection functions, a peak in the variance in the frequency of allele A_1 is evident at $x \simeq 0.5$ for $\beta = 3.5$.

DISCUSSION

With homogeneous selection, the results from this model are quite straightforward. The ratios among the genotypes become fixed everywhere in space, and no spatial cline in genotype frequencies can develop, since no spatial

structure is allowed. Genetic diversity can only persist if the fitness of the heterozygote is greater than the fitness of either homozygote – i.e. if there is overdominance.

The genetic dynamics with spatially inhomogeneous selection are more interesting. For a long generation time, with symmetric selection functions (i.e. no dominance), the shape of the time-averaged allele frequency cline is determined by the shapes of $\psi_{1,0}$ and $\psi_{3,0}$, and the variance in allele frequencies at the edges of the domain increases with increasing β. For a one-mode truncation with asymmetric selection functions (i.e. total dominance), the structure of the time-averaged allele frequency cline is again determined by the first eigenfunctions, but the average frequency of allele A_1 is lower than in the symmetric case. Also, in the asymmetric case, the average frequency of A_1 decreases with increasing β, and the variance in allele frequencies increases with increasing β.

Results with a short generation time have not been fully explored, but some of the implications of a short generation time can be seen in a comparison of a one- and a three-mode truncation. A three-mode truncation allows the time-averaged allele frequency cline to change shape, depending on the value of β. As in the model runs with a long generation time, the slope of the cline decreases with increasing β, and the variance in allele frequencies increases with increasing β. Because the three-mode case allows variability in the shape of the allele frequency cline, reaction-diffusion dynamics can result in unusual spatial structure in allele frequencies. With a three-mode truncation and a short generation time, the model behavior with asymmetric selection is very similar to that with symmetric selection.

Possibly the most interesting finding of the model is that, for long generation times, results with more than one mode are identical to those with a single mode. That is, for large T, the shape of the allele frequency cline is dictated entirely by the first eigenfunction of the solution to the selection-diffusion equation (1). This may have implications for real populations of organisms with long generation times, because it suggests a direct relationship between the shape of the selection functions themselves, and the shape of the resulting allele frequency cline.

Also of importance to natural populations is the implication that chaotic population dynamics can damp allele frequency clines for populations with short generation times. This finding suggests that to find steep clines in nature, one may have to sample the entire domain essentially instantaneously. To test the implications of this model for natural populations, one would need considerable data on the relevant parameters for the real population:

the diffusivity, κ, the generation time, T, and the death rate, μ. Fairly detailed information about the shape of the selection functions would also be necessary. Finally, only a population with discrete generations could be modeled.

Future Directions
An important question about this model is how its behavior differs from that of existing models (e.g. Keller, 1984), in which Hardy-Weinberg equilibrium is assumed *a priori*.

There are also several assumptions built into this model that should be corrected. First, the model currently assumes that density dependence affects the birth rate, but does not influence the death rate. By including a density dependent expression in the selection-diffusion equation, (1), (i.e., $\mu_i(x, \rho_1, \rho_2, \rho_3)$), one could explore the effects of more realistic density dependence. This addition would make the eigenfunction solution impossible, however. It would also be interesting to investigate the effect of allowing reproduction to occur over a short but finite time interval, rather than in a discrete pulse. Finally, the results described here pertain to a diffusionally dominated regime. Modifications are currently underway to allow investigation of the model's behavior with very slow rates of diffusion.

Application of this model to oceanographic questions would require expansion of the formulation to at least two spatial dimensions and addition of advective terms. With these additions, the model would have to be solved numerically. The genetics equations could then be included in a physical circulation model, to investigate the impact of ocean circulation patterns (of various scales) on genetic dynamics.

References

1. Hartl, D. L., and A. G. Clark, 1989. *Principles of Population Genetics, 2nd ed.* Sinauer Assoc., Inc., Sunderland, MA.

2. Keller, J. B., 1984. Genetic variability due to geographical inhomogeneity. J. Math. Biology **20**: 223-230.

3. Murray, J. D., 1993. *Mathematical Biology, 2nd ed.* Springer-Verlag, Berlin.

4. Nagylaki, T., 1975. Conditions for the existence of clines. Genetics **80**: 595-615.

5. Rasband, S.N., 1990. *Chaotic dynamics of nonlinear systems.* Wiley, New York.

6. Slatkin, M., 1973. Gene flow and selection in a cline. Genetics **75**: 733-756.

7. Webb, G.F., 1981. A genetics model with age dependence and spatial diffusion. In: *Differential equations and applications in ecology, epidemics, and population problems,* S.N. Busenberg and K.L. Cooke, editors, Academic Press, New York, pp. 29-40.

List of Figures

239

Figure 1

241

(a)

(b)

(c)

Figure 2

(a)

(b)

(c)

Figure 3

243

Figure 4

Figure 5

Figure 6

245

Figure 7

Figure 8

246

Figure 9

Figure 10

247

Figure 11

Figure 12
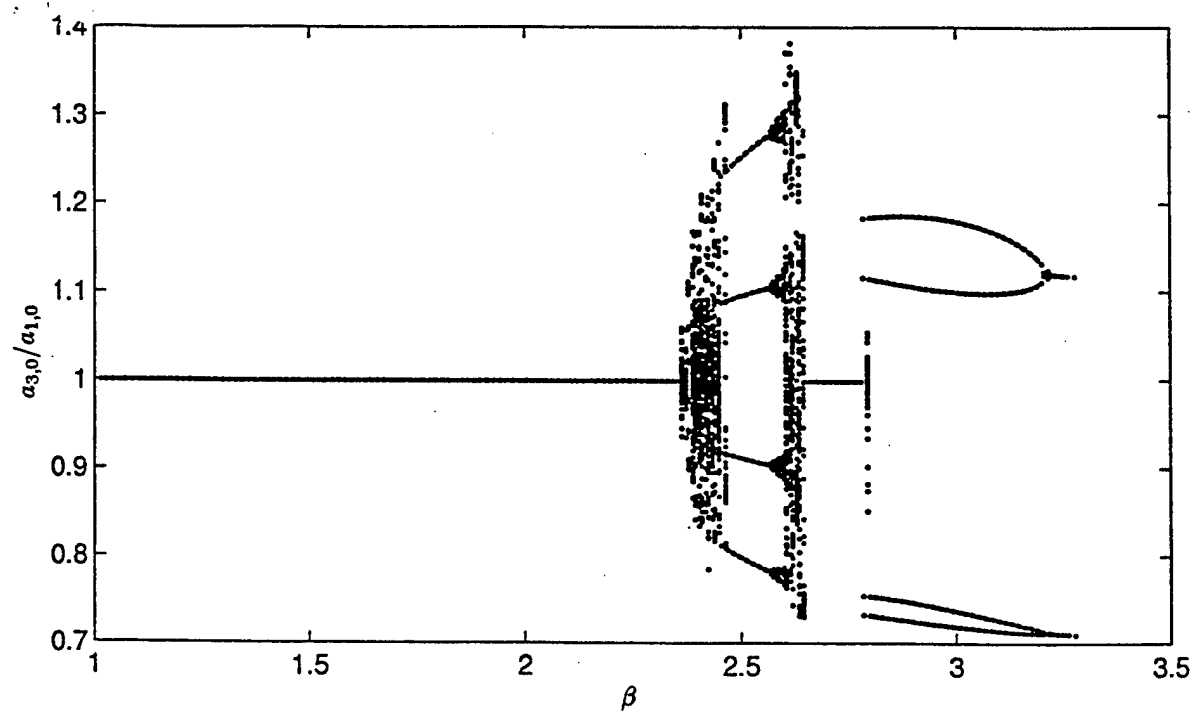
Figure 13
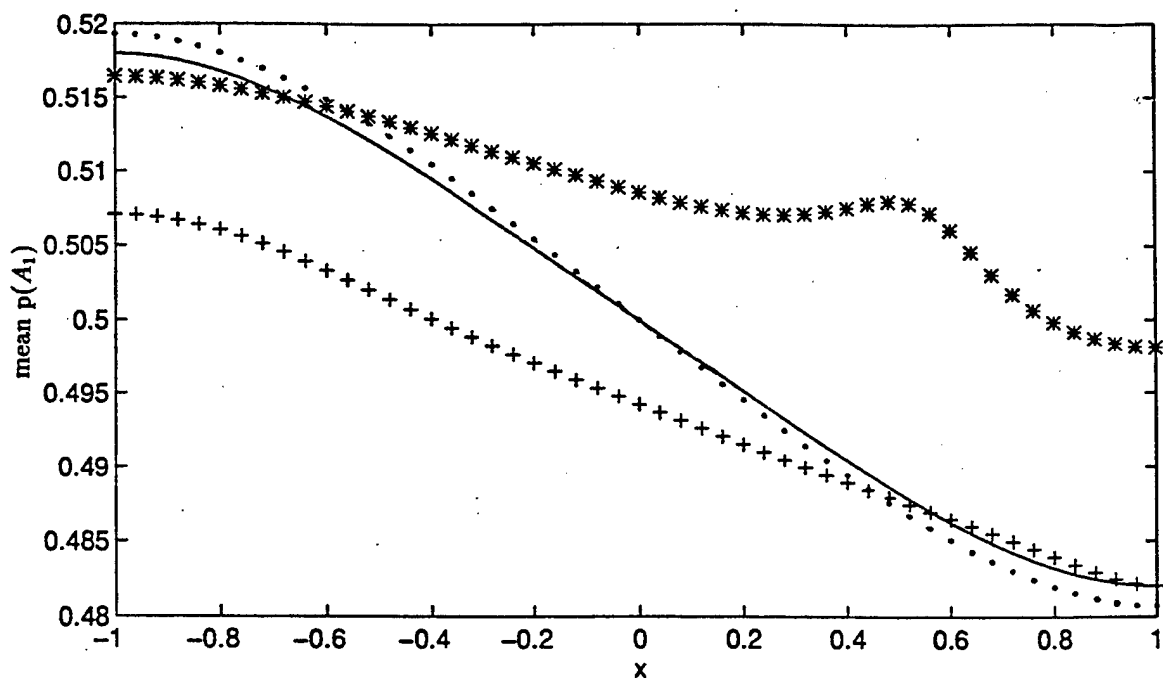
Figure 14

Figure 15

250

(a)

(b)

(c)

(d)

Figure 16

251
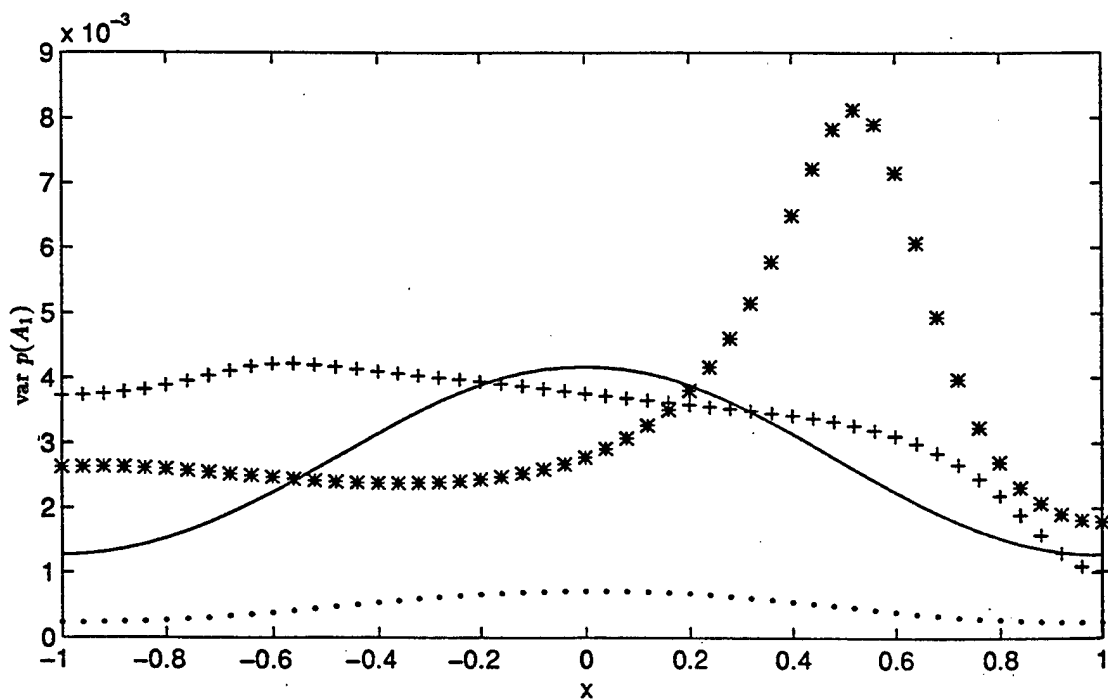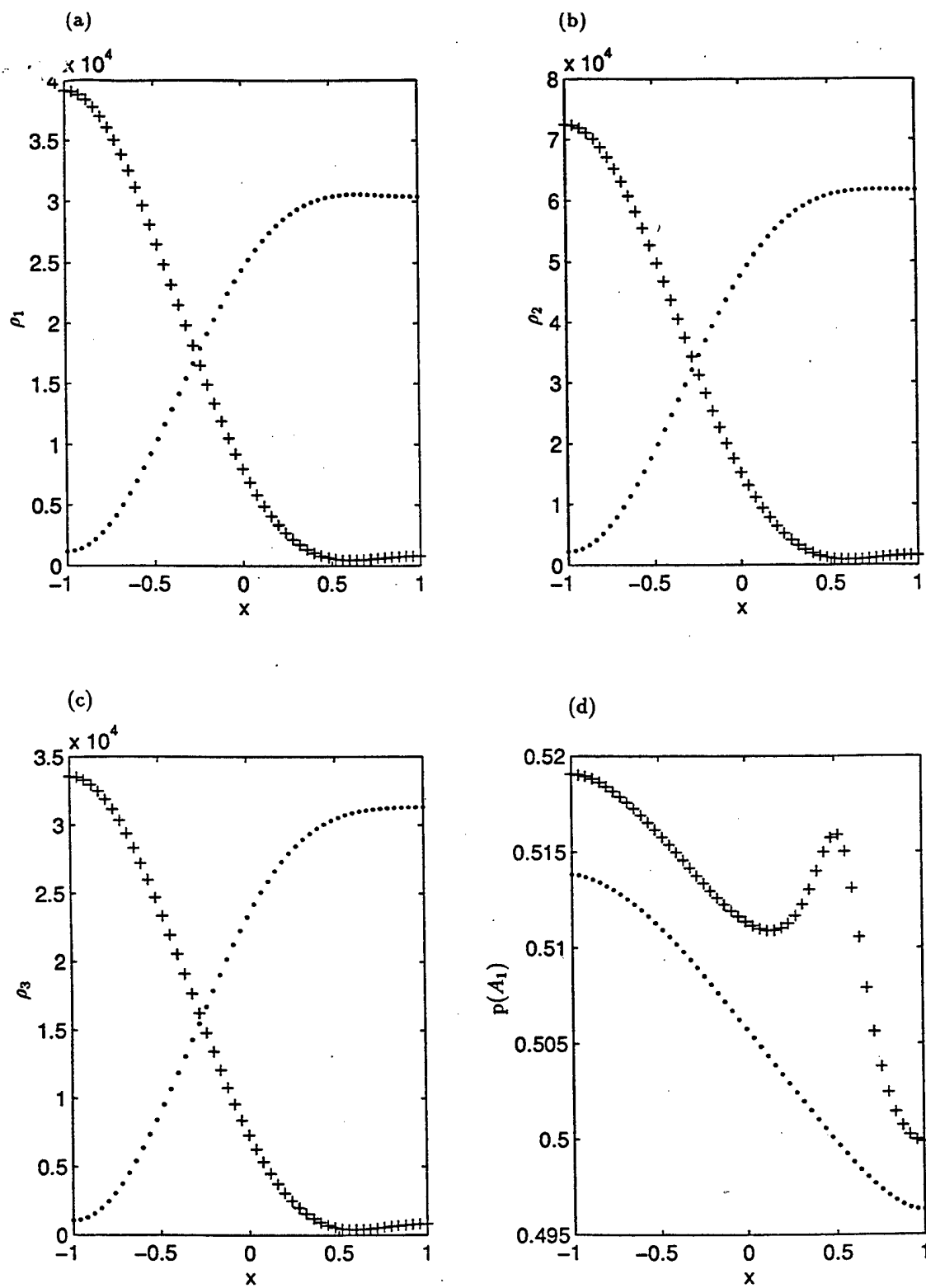
Preliminary Exploration of Three Upper Ocean Biology Models

J. Scott Stewart
Geophysical Fluid Dynamics Summer Program
Woods Hole Oceanographic Institute

Summer, 1994

1 Introduction

Life in the oceans plays an important role in modulation of carbon dioxide and other gas fluxes in the atmosphere. In situ measurement of such life and their effects are impractical on the large scales of interest in climate modelling. To determine the role of biology in physical processes, simple dynamical models which can be incorporated into somewhat more detailed physical models are necessary. In order to intelligently apply these models and interpret the results, however, it is necessary to determine the dynamical behavior of these equations in isolation. The subject of this project is to examine the dynamical behavior of three similar nitrogen-based biological models. The models are then subjected to simple forcing in total nitrogen and light intensity which roughly simulates seasonal cycles of these variables. It is interesting to determine if such an effect as the spring bloom can be reproduced by these models.

2 Model Descriptions

The biological models here consist of four coupled ordinary differential equations. The four compartments of the model Nutrients (N), phytoplankton (P), zooplankton (Z) and a detrital pool (D) are the four biological compartments of each model. Each is measured in terms of units of nitrogen and the total nitrogen of the ecosystem (N_0) is held constant. Figure 1 illustrates the relationship of these compartments to one another.

The governing equations for model 1 are

$$\dot{N} = -uIP\frac{N}{(K_s + N)} + eD + (1 - \alpha)gPZ$$

$$\dot{P} = uIP\frac{N}{(K_s + N)} - gPZ - sP$$

$$\dot{Z} = \alpha gPZ - dZ$$

$$\dot{D} = dZ + sP - eD$$

$$N_0 = N + P + Z + D \qquad (1)$$

representing linkages as shown in Figure 1. Note that since total nitrogen is conserved,
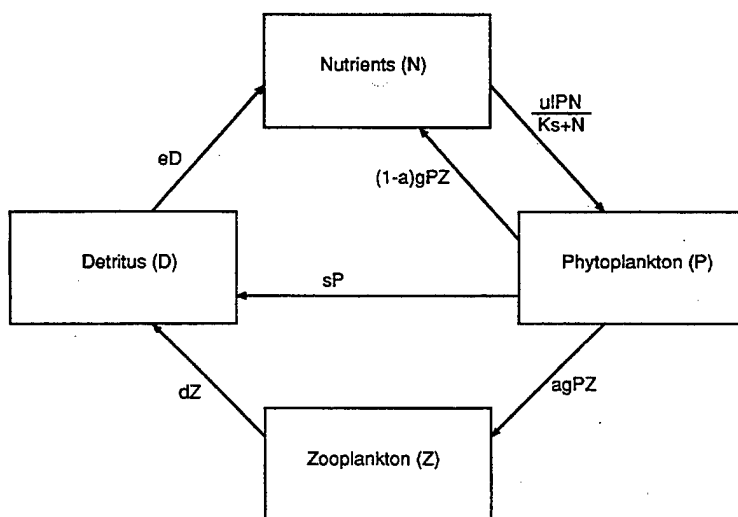
this set can be reduced to three independent equations. It is instructive, however, when analyzing the plots to show the value of nitrogen in each compartment, realizing that the total quantity is fixed. In these equations, u is the phytoplankton growth rate; I is the average incident photosynthetically active radiation; K_s is the nutrient half saturation constant; e is the detrital regeneration rate; α is the assimilation efficiency of the zooplankton; g is the zoolplankton growth rate; s is the photoplankton senescence (death due to aging) rate; and d is the zooplankton death rate.

The second and third models considered here differ in the treatment of the zooplankton uptake and zooplankton death rates respectively. The equations for these models are given in equations 2 and 3.

$$\dot{N} = -uIP\frac{N}{(K_s + N)} + eD + (1 - \alpha)gZ(P - Z)$$

$$\dot{P} = uIP\frac{N}{(K_s + N)} - gZ(P - Z) - sP$$

$$\dot{Z} = \alpha gZ(P - Z) - dZ$$

$$\dot{D} = dZ + sP - eD$$

$$N_0 = N + P + Z + D \qquad (2)$$

$$\dot{N} = -uIP\frac{N}{(K_s + N)} + eD + (1 - \alpha)gPZ$$

$$\dot{P} = uIP\frac{N}{(K_s + N)} - gPZ - sP$$

$$\dot{Z} = \alpha gPZ - dZ^2$$

$$\dot{D} = dZ^2 + sP - eD$$

$$N_0 = N + P + Z + D \qquad (3)$$

The second model differs from the first in that the grazing term includes a quadratic self-inhibition so that the zooplankton growth is logistic. The third model differs from the first in that the zooplankton death term is quadratic. This is suggested in Steele and Henderson (1992) as more accurately representing zooplanktonic competition. For this project, constant values are used for each of the coefficients. The values used are typical of copepod populations.

The analysis procedure consists of finding the stable states of each model and linearizing about these points to determine whether these points are linearly stable or unstable and, if stable, to determine the complex rate of relaxation back to the stable state.

3   Conditions for the Spring Bloom

As described by Sverdrup (Sverdrup, 1953), the spring bloom is set up when nutrient-rich waters are exposed to longer periods of daylight in the spring. High light, high nutrient environments are very favorable for phytoplankton growth and a rapid period of growth ensues. Eventually, however, much of the nutrients are used up and, despite the favorable light field, new production is limited. This signal is very evident in satellite images of ocean color, from which an estimate of chlorophyll in the euphotic zone may be obtained.

In order to understand the conditions for spring bloom, it is useful to examine the growth characteristics of a phytoplankton population in the mixed layer. The upper layer of the ocean may be considered to be well mixed to a certain depth called the mixed layer depth (MLD). The MLD is determined by a balance between stratification effects such as surface

warming and precipitation, and destratification effects such as radiative cooling, evapora-
tion, and wind stress. During summer days, when total light intensity is higher, the mixed
layer tends to be shallower than during winter. Phytoplankton are generally believed to be
incapable of controlling their vertical position in the water column and are therefore, like
physical constituents, well mixed in the mixed layer. Over long enough time scales, they can
be considered to spend an equal portion of their time at all depths in the mixed layer. This
leads to the assumption that the average light exposure of the population is the average light
exposure of the mixed layer. Sverdrup assumed that the phytoplankton respiration is the
same at any depth. Respiration here means the energy used up by the phyoplankton simply
by existing in their environment. It includes both normal metabolic processes and the effects
of grazing. The production depends upon the light intensity which decreases exponentially
with depth. Figure 2 illustrates the respiration and production rates as a function of depth.
The respiration rate is shown as the rate at which the biomass would decrease if no pro-

Figure 1: Effects of Production and Respiration on Phytoplanton Biomass
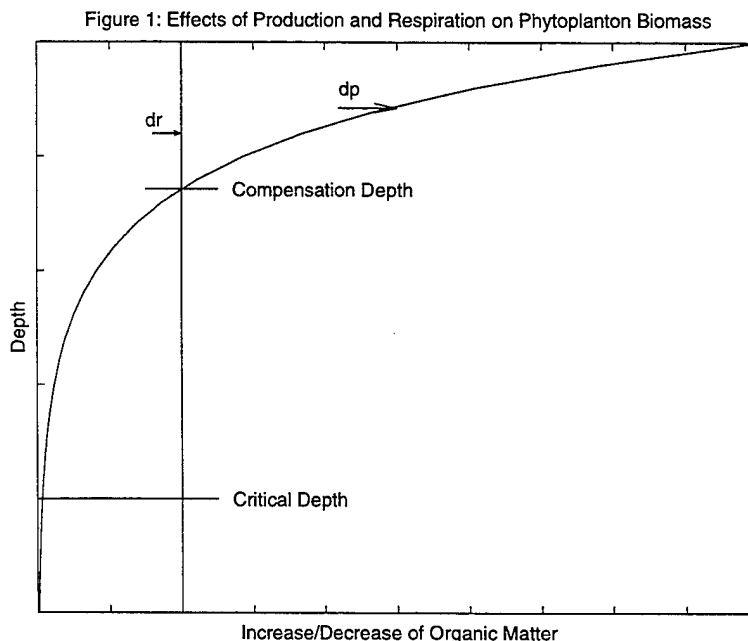
duction occurred. The net rate for the mixed layer is the depth-averaged value of the area
to the left of the *dr* curve above the mixed layer depth in Figure 2. The production rate
is the rate at which the biomass would increase if there were no respiration. For a given
MLD, this corresponds to the area above the *dp* curve and above the MLD in Figure 2. The
compensation depth is the depth at which the production rate equals the respiration rate.
The critical depth is the depth of the mixed layer at which the total respiration is exactly
balanced by the total production. The phytoplankton biomass will increase if the mixed
layer is shallower than the critical depth and will decrease if the mixed layer is deeper.

During the spring bloom, the mixed layer shoals due to surface heating by increased light
as the days lengthen and the sun rises in the sky. When the mixed layer is shallower

255

than the critical depth, the phytoplankton take advantage of the high nutrient environment established during the winter by both remineralization of non-nutrient nitrogen in the water and mixing with deeper, higher nutrient waters when the MLD was greater.

Sverdrup (1953) presents data from Weather Ship "M" in the Norwegian Sea from March to May, 1949. This is reproduced in Figure 3. The hashed region is the estimated critical

Figure 2. Results of observations at Weather Ship "M" (66°N., 2°E. Gr.). The symbols are explained in the graph, where the following abbreviations have been used:— Dia, Diatomaceae; Coc, Coccolithophoridae; Dif, Dinoflagellatae; Nau, Nauplii; and Cop, Copepods.
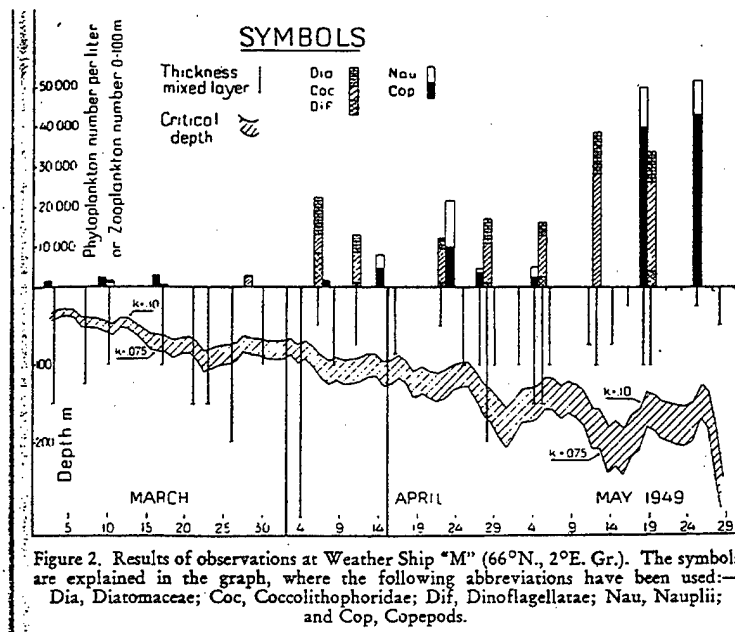
depth. This is based on water clarity, insolation and uncertainties about the light extinction coefficient. The biology data are classified by phytoplankton (diatoms, coccolithophores and dinoflagellates) and zooplankton (nauplii and copepods). The solid lines indicate measurements of the mixed layer depth. Until April, it is clear that the mixed layer is generally deeper than the critical depth, and the biological signal is low. At the end of April and the beginning of May, however, the mixed layer has shoaled sufficiently and the plankton counts increase. By late May, the bloom is over; the phytoplankton population is again low, having been largely eaten by zooplankton. The only departure from this general scheme is about two weeks of data in early April in which a shallow mixed layer and high phytoplankton counts are recorded. This is likely due to advection of a foreign water mass.

4 Analysis of the Model Equations

Expressions for the fixed points of the first set of equations were determined analytically. The derivatives were set to zero and the values for N, P, Z, and D were calculated algebraically. Solutions were obtained for the steady-state values of N, P, Z and D, but the full analytical

256

solutions in terms of the coefficients of the equations are rather tediously obtained. For the first and second equation sets, this portion of the derivation was carried out by the symbolic processing program Mathematica. The solutions are recorded in the appendix. For the third model, the computation (which involves symbolically solving for a few square roots) is too excessive. For this problem, the steady states are determined as needed using a root finding program.

The first and second steady states correspond to coexisting phytoplankton and zooplankton populations and are given by:

$$
\begin{aligned}
\hat{P} &= \frac{d}{\alpha g} \\
\hat{Z} &= \frac{uI\hat{N}^* - s}{g} \\
\hat{D} &= \frac{\frac{d}{g}uI\hat{N}^* - s + s\hat{P}}{e} \\
\hat{N} &= \frac{K_s\hat{N}^*}{1 - \hat{N}^*} \\
\hat{N}^* &= \frac{\hat{N}}{K_s + \hat{N}} \\
&= \frac{\hat{P} + \hat{Z} + \hat{D} - N_0}{\hat{P} + \hat{Z} + \hat{D} - N_0 - K_s} \\
&= \frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \\
\text{where} \quad A &= \frac{uI}{g} + \frac{duI}{eg} \\
B &= -C - A - K_s \\
C &= N_0 - \frac{d}{\alpha g} + \frac{s}{g} + \frac{ds}{\alpha g} - \frac{sd}{\alpha ge}
\end{aligned}
\tag{4}
$$

where $\hat{N}^*$ is an intermediate solution variable used to simplify the notation.

The third steady state has no zooplankton population and is given by:

$$
\begin{aligned}
\hat{P} &= \frac{N_0 - \hat{N}^*}{1 + \frac{s}{e}} \\
\hat{N} &= \frac{K_s\hat{N}^*}{1 - \hat{N}^*} \\
\hat{Z} &= 0
\end{aligned}
$$

257

$$\hat{D} = \frac{s}{e}\frac{N_0 - \frac{K_s s}{uI - s}}{1 + \frac{s}{e}}$$

$$\hat{N}^* = \frac{s}{uI} \tag{5}$$

The fourth steady state simply contains nutrients:

$$\hat{N} = N_0$$

$$\hat{P} = \hat{Z} = \hat{D} = 0 \tag{6}$$

For the second model, a similar set of steady states are solved for:

$$\hat{N} = N_0 - \hat{P} - \hat{Z} - \hat{D}$$

$$\hat{P} = \frac{d}{\alpha g^2}uI\hat{N}^* - s$$

$$\hat{Z} = \frac{uI\hat{N}^* - s}{g}$$

$$\hat{D} = \frac{d\hat{Z}}{e}\hat{Z} + \frac{s}{\alpha g}$$

$$\hat{N}^* = \frac{\hat{N}}{K_s + \hat{N}} \tag{7}$$

which results in a lengthy square root (not recorded here) similar to that derived from model 1. The third steady state is:

$$\hat{N} = \frac{sK_s}{uI - s}$$

$$\hat{P} = \frac{N_0 - \frac{sK_s}{uI - s}}{1 + \frac{s}{e}}$$

$$\hat{Z} = 0$$

$$\hat{D} = \frac{s}{e}\frac{N_0 - \frac{sK_s}{uI - s}}{1 + \frac{s}{e}} \tag{8}$$

and the fourth steady state is:

$$\hat{N} = N_0$$

$$\hat{P} = \hat{Z} = \hat{D} = 0 \tag{9}$$

Both the first and second models have 4 fixed points, whereas the third model, with the quadratic zooplankton competition term, has 5 fixed points. The first three steady states are too complicated to solve for analytically, and so are computed using a root finding routine from the following:

$$\hat{P} = \frac{-g\hat{Z}^2}{uI\hat{N}^* - s - g\hat{Z}}$$

$$\hat{Z} = \frac{\frac{sd}{\alpha g} - \frac{duI\hat{N}^*}{\alpha g}}{uI\hat{N}^* - s - \frac{d}{\alpha}}$$

$$\hat{D} = \frac{s\hat{P}}{e}$$

$$\hat{N} : A\hat{N}^3 + B\hat{N}^2 + C\hat{N} + D = 0$$

$$\text{where } A = -g^2$$

$$B = -2g^2K_s + g^2N_0 + \frac{ds}{\alpha} + gs - \frac{ds^2}{e} + \frac{ds^2}{\alpha e}$$
$$- \frac{duI}{\alpha} - guI + \frac{2dIsu}{e} - \frac{dIsu}{\alpha e} - \frac{dI^2u^2}{e}$$

$$C = -g^2K_s^2 + 2g^2K_sN_0 + \frac{2dK_ss}{\alpha} + 2gK_ss - \frac{2dK_ss^2}{e}$$
$$+ \frac{2dK_ss^2}{\alpha e} - \frac{dIK_su}{\alpha} - gIK_su$$
$$+ \frac{2dIK_ssu}{e} - \frac{dIK_ssu}{\alpha e}$$

$$D = g^2 + K_s^2N_0 + \frac{dK_s^2s}{\alpha I} + gK_s^2s - \frac{dK_s^2s^2}{e} + \frac{dk_s^2s^2}{\alpha e} \qquad (10)$$

The fourth steady state is:

$$\hat{N} = \frac{sK_s}{uI - s}$$

$$\hat{P} = \frac{N_0 - \hat{N}}{1 + \frac{s}{e}}$$

$$\hat{Z} = 0$$

$$\hat{D} = \frac{s\hat{P}}{e} \qquad (11)$$

The final steady state is the familiar all-nutrient state:

$$\hat{N} = N_0$$
$$\hat{P} = \hat{Z} = \hat{D} = 0 \qquad (12)$$

The important thing to notice about the steady state calculations is that in each of the models, there are four or five stable points. One is an all nutrient state, in which case there is no biology. This indicates an environment in which the phytoplankton are not in a sufficiently suitable light and/or nutrient environment, cannot grow and therefore decay away at their senescence rate. Another steady-state is a zero- zooplankton state, in which case there is no grazing of the plants. While it is unlikely that zooplankton populations would ever die out in the real world, populations may tend toward this state, and it is retained in the analysis. A third is an equilibrium state between phyto- and zooplankters. (Fourth and fifth stable points consist of negative values for some of the compartments and are unrealizable.) Assuming that the biological constants such as growth rates, etc remain the same despite physical forcing, the only externally forced parameters are the total nitrogen, N_0 and the average light intensity, I.

For this first study, the models are assumed to represent the state of a bulk mixed layer. The light intensity is set to the average light level of the layer and the total nitrogen is the nitrogen in the layer. In order to understand the behavior of these equations, it is useful to get a feel for the behavior in the state-space of physical forcing: N_0 and I. The modulation of light intensity may be likened to seasonal light level cycles. The changing N_0 to the use and subsequent sinking of nitrogen as the summer wears on and the mixed layer becomes depleted of nitrogen. In future studies, these changes will be determined by a mixed layer code, but in this first look at equation behavior, we will apply a forcing in N_0 and I.

First, however, it is useful to plot the stability of each physically realizable steady state versus the environmental conditions. It turns out that for a given forcing in N_0 and I, only one of the steady states is stable, and therefore that solutions will tend to that state. It is also useful to examine the complex decay coefficient to each of these states since both highly damped and highly oscillatory paths to the steady state are possible. Here, however, only a linear stability analysis about the fixed points is considered, limiting the usefulness of the decay coefficient at points not near a stable curve.

The appropriate stability matrix is computed manually and is illustrated for each of the three models. Figure 4 illustrates the stability regimes for the first model. Clearly, if both total nitrogen and light intensity are high, both phyto- and zooplankton may coexist in some steady-state. If either variable is too low, life dies out. In between, however, there is a region where zooplankton will die out, but phytoplankton will continue to persist.

For the second model, a similar plot can be generated. The steady-state boundaries are similar and the same three states: coexistence, no-zooplankton and all-nutrients, are observed.

The third model, however, is somewhat different. First, as illustrated here, the N_0-I space explored is larger. This is to show that the three roots of the coexistence polynomial are all stable for different regimes. The interesting feature, however, is that the shape of the steady-state nutrient and phytoplankton populations does not betray the location of these basin-

Steady-state basins of attraction—Model 1
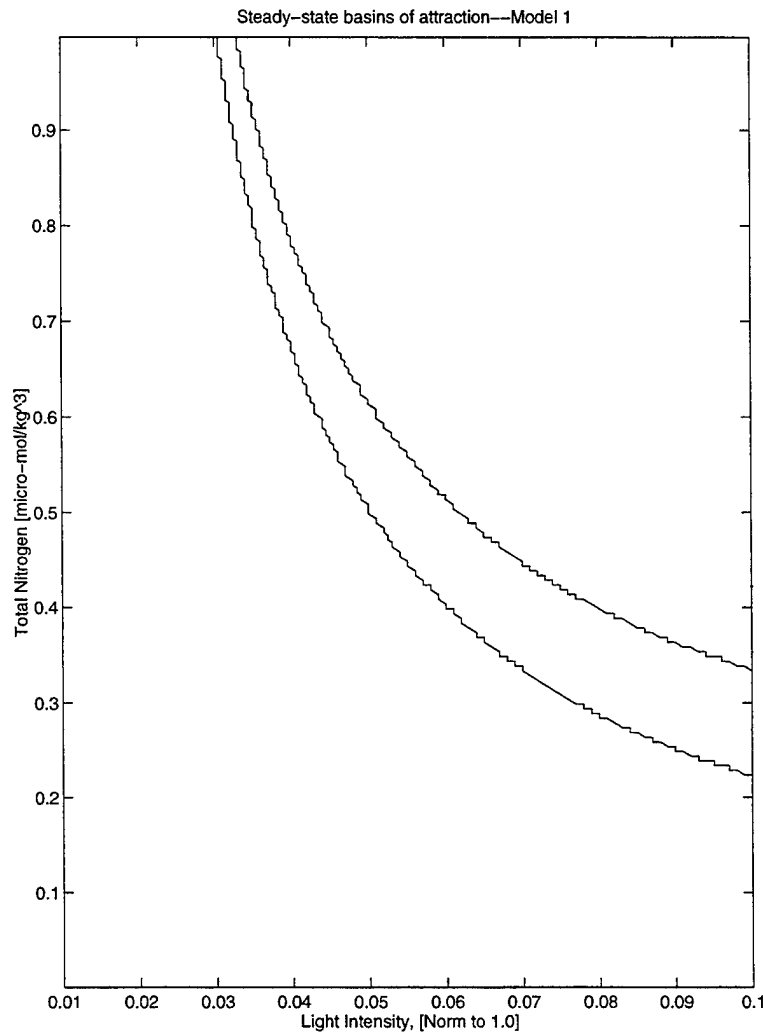
boundaries. That is to say, only one of the three roots yields physically realizable populations, but in the limit as the external forcing approaches the stability-range boundaries, the steady-state populations approach the same values. A comparison may be made by examining Figure 7, wherein the steady-state nutrient and phytoplankton populations are presented as a fraction of total nitrate, N_0 available.

Examination of Figure 7 indicates that there is little difference between the first two models. Steady-state nutrient and phytoplankton levels are almost identical for any given nutrient/light forcing values. The third model yields distinctly different results. Most noticeable is that the no-zooplankton attractor range is not bounded by a curve similar to the all-nutrient attractor range. Rather, a definite amount of total nitrogen is necessary in the system in order for zooplankton not to die out. Thus the inclusion of a quadratic competition in the zooplankton growth rate equations changes the behavior of the equations considerably.

Steady–state basins of attraction––Model 2
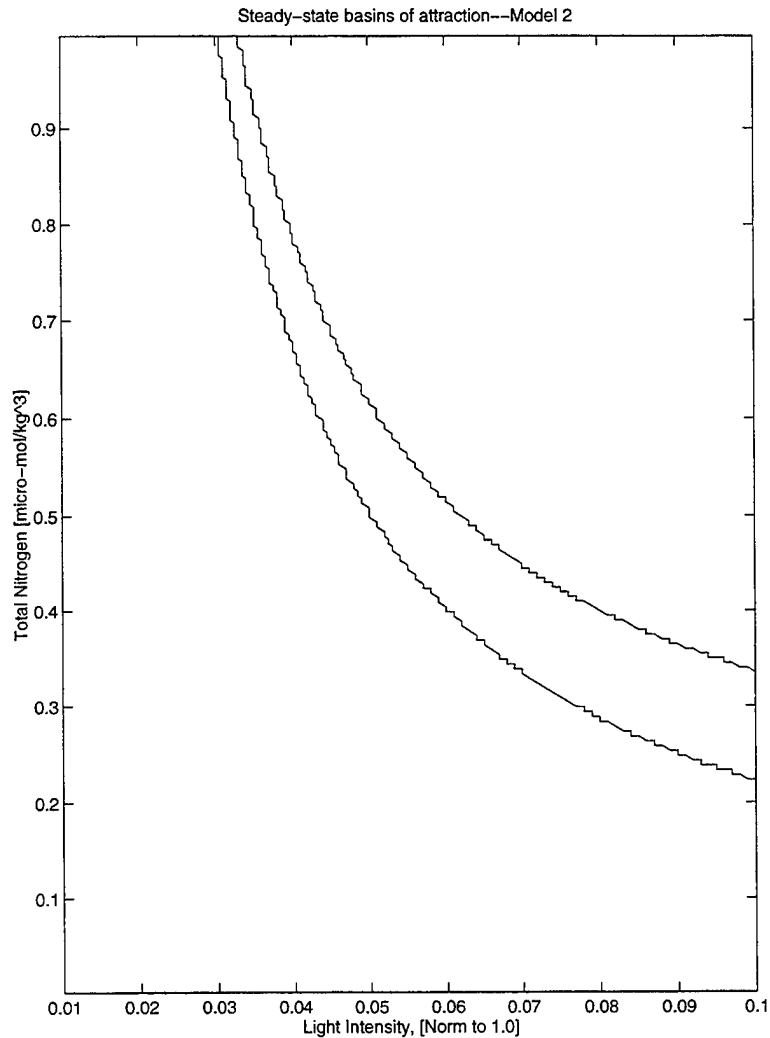
The next step in this line of inquiry will be the analysis of these types of equations in simple mixed layer models of upper ocean mixing. Such a study differs from the present case in that both the light and nutrient forcing will be different for different levels in a mixed layer model and these values will be time-varying. Additional considerations are several. It will be necessary to consider how to treat the vertical mixing of the different biological components, including independent movement rates of phytoplankton and zooplankton and the perhaps different sinking rates of detrital matter. Also, the response of phytoplankton to a time-varying light field must be considered. Is their productivity proportional to the average light obtained over a time period, or is the response more nonlinear. Also, how much does the spatially nonuniform turbulent mixing of the phytoplankters affect the bulk productivity. It will be interesting to consider these questions, especially if data sets which can serve to validate these modelling approaches can be obtained.

262

Steady–state basins of attraction—Model 3
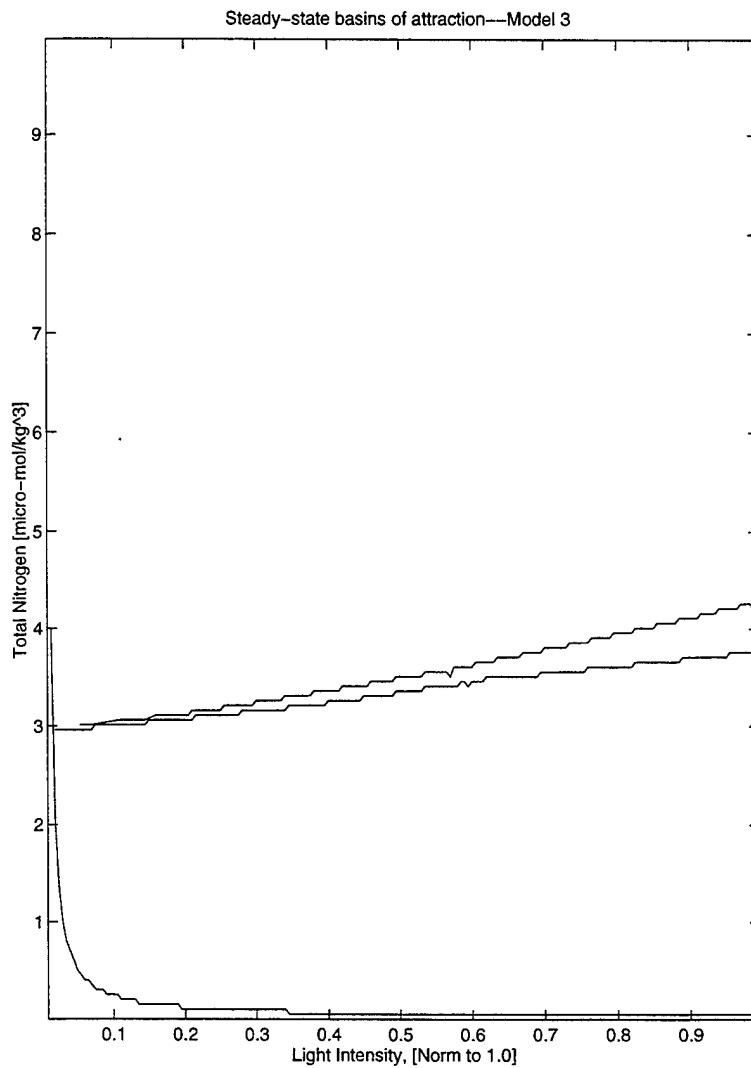
5   References

Sverdrup, H.U., On Conditions for the Vernal Blooming of Phytoplankton, *Phytoplankton*, 1953, 287-295.

Steele, J.H. and E.W. Henderson, 1992: The role of predation in plankton models. *J. Plankton Res.*, **114**, 147-172.

263

Steady-state Nutrient [%age of No]--Model 1
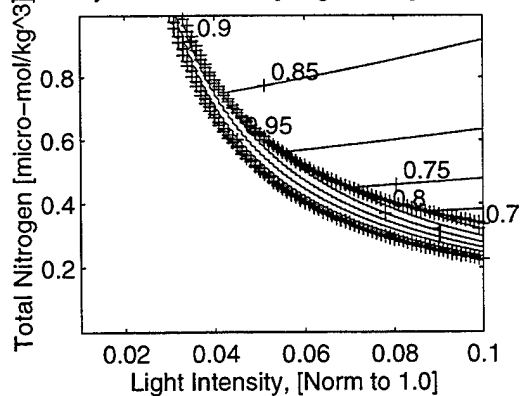
Steady-state Phytoplankton [%age of No]--Model 1
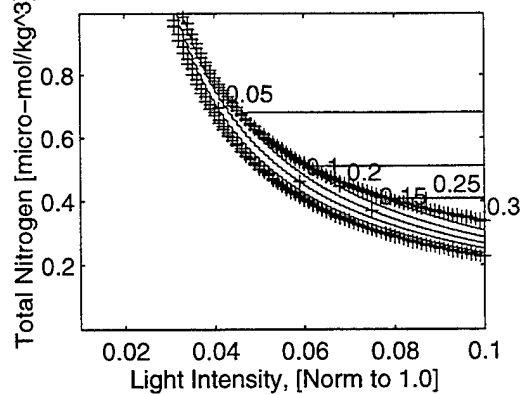
Steady-state Nutrient [%age of No]--Model 2
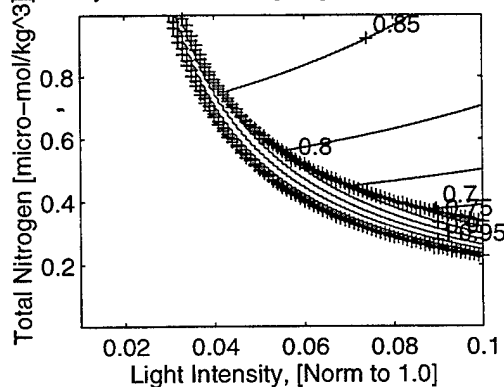
Steady-state Phytoplankton [%age of No]--Model 2
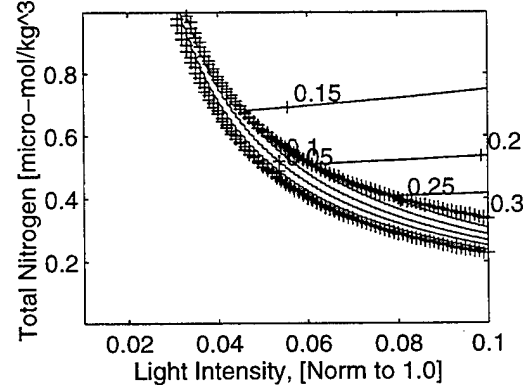
Steady-state Nutrient [%age of No]--Model 3
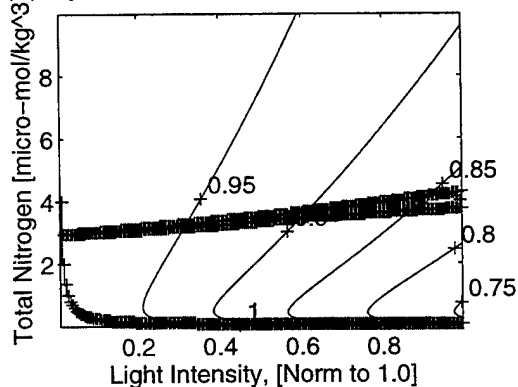
Steady-state Phytoplankton [%age of No]--Model 3

264
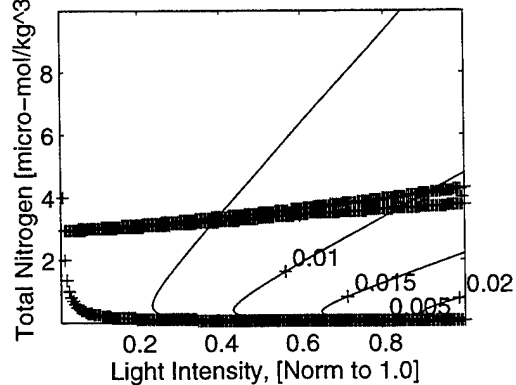
The Birds and Bees

- A Mathematical Model of Speciation

M A Bees

Department of Applied Mathematical Studies

University of Leeds

Advised by E A Spiegel

Department of Astronomy

Columbia University

New York NY 10027

December 7, 1994

I model and analyse the interaction between a species and the environmental carrying capacity for a continuous array of a species' distinguishing features. I find that gradual change interspersed with rapid separation into two species is the norm. The model shows aspects of both gradualism and the "punctuated equilibria" of Eldredge & Gould (1972).

1 Introduction and review

Speciation is the process of a single species breaking up into two distinct species with no crossbreeding. Evolutionists suggest there are two main views of how this occurs (see figure 1):

- Gradualism - the neo-Darwinian view that natural selection causes species to gradually evolve apart (1930's).

- Punctuated Equilibria - little change over long periods of time followed by abrupt speciation caused by a faster mechanism than just natural selection (Eldredge & Gould, 1972).
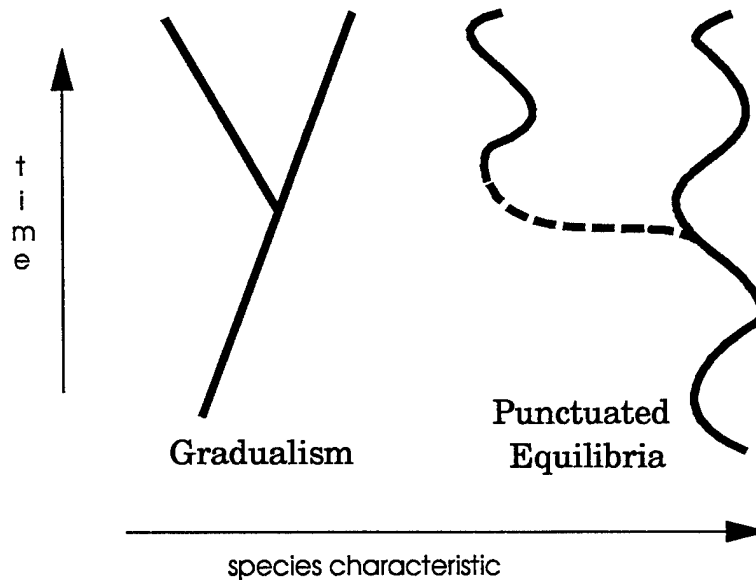
Figure 1: The two views on speciation

There are many examples where authors support one or other of these views and attempt to fit experimental data to them. Unfortunately, a general

consensus has not been reached and none of these attempts has proved convincing to the opposition. As an example of the ambiguity of the data see figure (2) where two independent authors claim that it is evidence for their own favorite view.

The aim of this investigation is to form a consistent model of evolution so that both of these views are catered for. They will be obtained as solutions for different parameter values of the same underlying mathematical formulation. It can be argued that a mechanism for causing speciation is geographical isolation. This is where a species is split and the two halves are now free to gradually evolve according to their independent environments. However this is not the only form of genetic isolation. Species can be isolated by overcrowding in environmental niches, as I shall show later. Geographical isolation is not necessarily evident in the sea or on a small island, but speciation still occurs. For the present I will ignore spatial effects. The environment must have an effect but how? Do we prescribe the environment or do we allow it to vary in relation to a species taking advantage of it? In this paper I opt for coupled environment-species equations.

In the absence of anything abusing the environment it should be able to grow to full capacity, so I arbitrarily choose logistic growth. Further, I say that the presence of a particular species using the environment will reduce the carrying capacity in a linear fashion. Now I will consider the species. Each species is treated as a collection of purely selfish individuals, as advocated by Dawkins (1989). There are two paths we can follow here. Either we model the genetics as in numerous papers on clines in allele space (Voronka & Keller (1975), Ewens (1969), O'Brien (1985) and Spencer & Barakat (1992) - to name a few), or we model the phenotype. I opt for the latter model of phenotype, thus disregarding a species' genetic makeup. I will extend the phenotypic descriptor to a more general characteristic.

H. wortmani   H. Mentalis

H. lysitensis

H. powellianus

H. minor   H. miticulus

H. simplex

H. latidens

H. loomisi

Lost cabin

Lysite

Gray bull

1600
1400
1200
1000
800
600
400
200

?

.80   .90   1.00   1.10   1.20   1.30

LOG (L × W) OF M_1
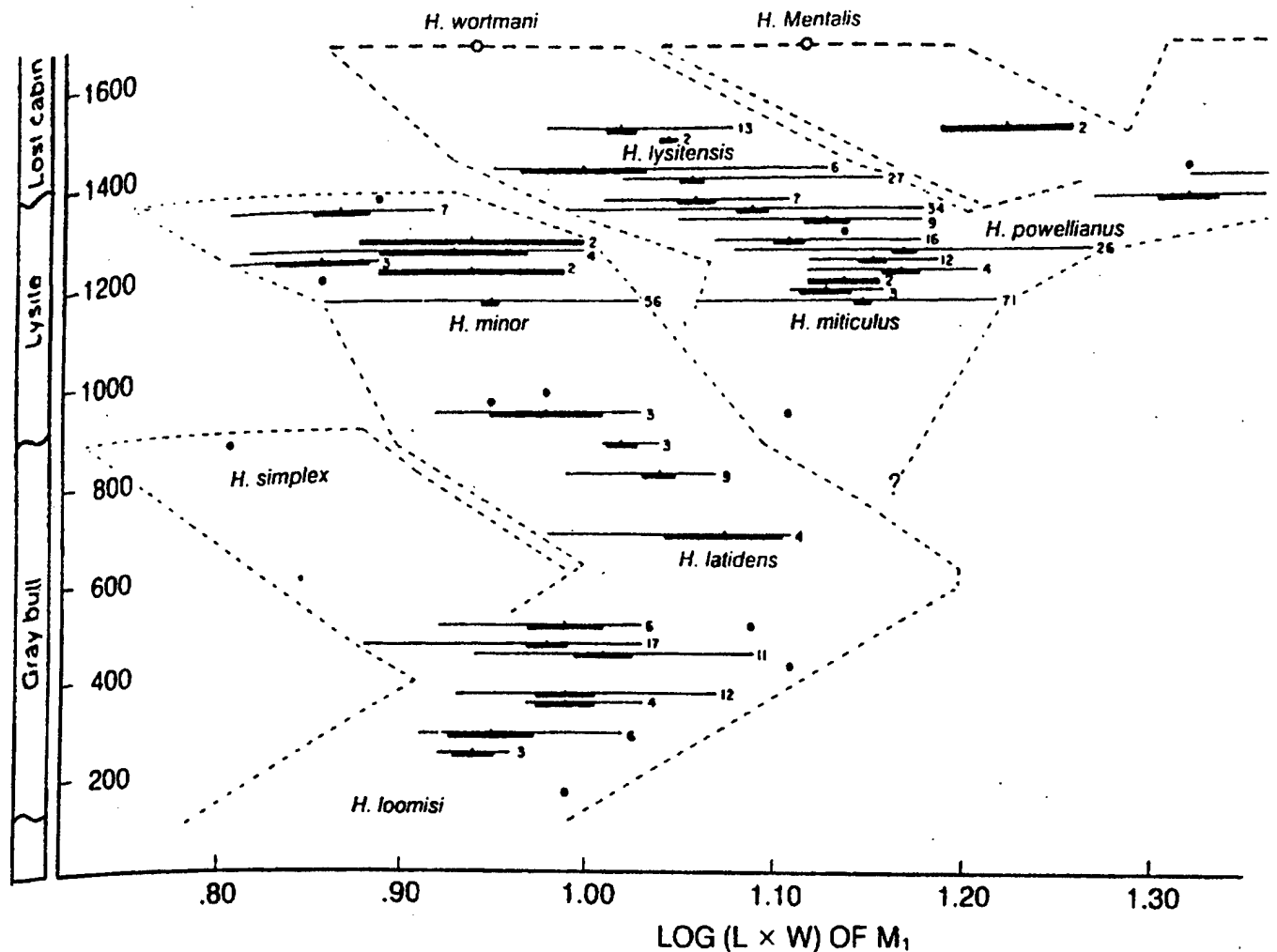
Figure 2: An example of conflicting interpretations of data from the fossil record .

A particular animal can have a huge number of identifying characteristics that distinguish it from other species. These characteristics could be anything from the animals' habitat, mating dance, a length or mass, or just appearance. Weiner (1994) presents details of research into the beak shapes of finches on the Galápagos Islands and compares qualitative and quantitative field work to molecular studies on selection. He suggests that naturalists can measure selection as functions of behaviour whereas laboratory work fails to find any other than neutral selective adaptation. To simplify matters I will consider only one independent characteristic. The environmental capacity for holding a particular species will have a comparable distinguishing fea-

ture. (See Roughgarden (1972) for a discussion and experimental data). For example the species could be giraffes with characteristic height α, and the environment could be trees with leaves at a height α. Alternatively, I could consider humming birds that eat the nectar of a particular flower and hence require the correct beak size and shape.

To model natural selection I need, at least, to include mating, a death term and some form of mutation (or random genetic diffusion) in characteristic space. I shall say that the offspring will only survive if the environment can support them. The mating term must be able to support a small amount of breeding between two animals of slightly different aspect but a larger difference in aspect would constitute cross-breeding between species and should not be allowed. For this reason an integral representation of the process has been used with kernels that represent probability distributions that two individuals will mate and produce a particular offspring. I will assume that two species will mate with probability as a function of the distance between them in characteristic space. Also, I shall assume that if two species mate then they will produce an offspring with a characteristic probability density function centered on the average of its parents characteristics (as in Roughgarden (1972) and utilized by Levin & Segel (1982), but see Slatkin & Lande (1976) for a possible improvement).

2 The model

Consider a species $f(\alpha, t)$ with characteristic α and an environment $e(\alpha, t)$ with niche characteristic α. The equations described above are:

$$\partial_t e = se(K - e) - aef, \tag{1}$$

$$\partial_t f = re \iint W(\alpha_1, \alpha_2) B\left(\left|\frac{\alpha_1 + \alpha_2}{2} - \alpha\right|\right) R(|\alpha_1 - \alpha_2|) f(\alpha_1) f(\alpha_2) d\alpha_1 d\alpha_2$$

$$- \ vf(\alpha) + \int m(|\alpha_1 - \alpha|)f(\alpha_1)d\alpha_1 - \int M(|\alpha - \alpha_2|)f(\alpha)d\alpha_2, \qquad (2)$$

where s is the environmental growth rate, K is the environmental capacity, a is the abuse of the environment by a species, r is the species growth rate, v is the species death rate, B is the normalized birth distribution, R is the normalized relationship distribution, M is the mutation distribution and W is a weighting function.

W represents physical constraints on reproduction. For instance, W could limit reproduction due to choosing one and only one mate in their lifetime (eg humans - sometimes) or restrictions due to maximum pregnancies per year/lifetime. For convenience I shall investigate circumstances where $W = 1$, which corresponds to animals such as plankton or any animal that can lay as many eggs as it cares to, so long as they can come into contact with a viable mate. In this way the spatial dependence could come into the problem. For instance, in the sea, spatially coming into contact with a viable mate could present considerable difficulties and an animals reproductive efficiency must surely increase with concentration of potential mates, e.g. plankton. Of course this reproductivity is limited by the environment, such as nutrients and light, but otherwise, over sufficient time, the organism is free to breed at will. Hence, I can justify my assumptions in the mating term by the inclusion of the feedback mechanism of environmental overcrowding. This is similar to that seen with weeds, where they will rapidly increase in number if there are many similar weeds around, but will be limited in their reproductivity if there are too many weeds using the same resource. In the absence of any reproduction, the population exponentially decays due to the death term.

The mutation is modelled by first considering mutations from characteristic α to α_1 and then from α_2 back to α. If the kernel M is symmetric, which means it will mutate in either direction with equal probability, and it decays exponentially fast then this term can be expanded (Taylor series or steepest

descents) to the usual second order derivative representing diffusion, plus higher terms,

$$m_2 f^{II} + m_4 f^{IV} + m_6 f^{VI} + ..., \tag{3}$$

where the m's are constants depending on the symmetric moments of M. This is similar to the classical diffusion term plus higher orders. The actual size of the mutation term is unimportant as I can scale the time, characteristic and species variables to balance the terms at will. That is to say, there is a significant difference between very small diffusion and no diffusion. Scaling the environment by K, time by $\frac{K}{s}$, α by $\frac{Ks}{M}$, the species by $\frac{M^2}{rK^2s}$ and defining $p = \frac{aM^2}{K^3rs^2}$ and $\mathcal{D} = \frac{v}{Ks}$ then I have the non-dimensional set of equations

$$\partial_t e = e(1 - e) - pef, \tag{4}$$

$$\begin{aligned}
\partial_t f &= e \iint W(\alpha_1, \alpha_2) B\left(\left|\frac{\alpha_1 + \alpha_2}{2} - \alpha\right|\right) R(|\alpha_1 - \alpha_2|) f(\alpha_1) f(\alpha_2) d\alpha_1 d\alpha_2 \\
&\quad - \mathcal{D}f(\alpha) + \int m(|\alpha_1 - \alpha|) f(\alpha_1) d\alpha_1 - \int m(|\alpha - \alpha_2|) f(\alpha) d\alpha_2, \tag{5}
\end{aligned}$$

where m is the normalised mutation term and the σ's have all been scaled with the characteristic space scale.

I should note here that the term involving p could be replaced with an integral over characteristic space such as

$$e \int q(|\alpha - \alpha_1|) f(\alpha_1) d\alpha_1, \tag{6}$$

where q is a distribution of species eating the environment at a particular α. This could be a better represention of reality, but for simplicity I shall only consider the case $q(z) = p\delta(z)$, where p is a constant.

3 A First Approach

This section documents the first attempt at understanding the model. However, it failed as it did not fully decribe the behaviour of the solutions. There

271

are four homogeneous equilibrium points for equations (4) and (5):

$$(e, f) = (0, 0), (1, 0), \left(\frac{1}{2} \pm \frac{\sqrt{1 - 4\mathcal{D}p}}{2}, \frac{1}{2p} \mp \frac{\sqrt{1 - 4\mathcal{D}p}}{2p} \right).$$ (7)

Consider an oscillatory perturbation to the non-zero homogeneous equilibrium solutions.

$$\begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix}_0 + \epsilon e^{\sigma t + ik\alpha} \begin{pmatrix} e \\ f \end{pmatrix}_1.$$ (8)

Then to $O(\epsilon)$

$$(\sigma - 1 + 2e_0 + pf_0) \left(\sigma - 2f_0 e_0 \bar{R} \left(\frac{k}{2} \right) \bar{B}(k) - \bar{m}(k) + \bar{m}(0) + \mathcal{D} \right) = -f_0^2 e_0 p,$$ (9)

where a bar denotes Fourier transform. Choosing normal distributions for R, B and m we get that

$$\sup_k \sigma(k) = \sigma(0).$$ (10)

Hence, consider $k = k_{max} = 0$. Then

$$\sigma(0) = \frac{\mathcal{D} - e_0}{2} \pm \frac{\sqrt{(e_0 + \mathcal{D})^2 - 4p\mathcal{D}f_0}}{2}.$$ (11)

For a Hopf bifurcation we require

1. $\mathcal{D} = e_0$

2. $(e_0 + \mathcal{D})^2 < 4p\mathcal{D}f_0$.

This gives us the the behaviour of the system in parameter space as in figure (3). Around the equilibrium point that exibits the Hopf bifurcation and knowing that $k_{max} = 0$, we can consider a solution

$$\begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix}_0 + \epsilon \int e^{i\omega t} \mathbf{a}(k, \omega) dk$$ (12)

272
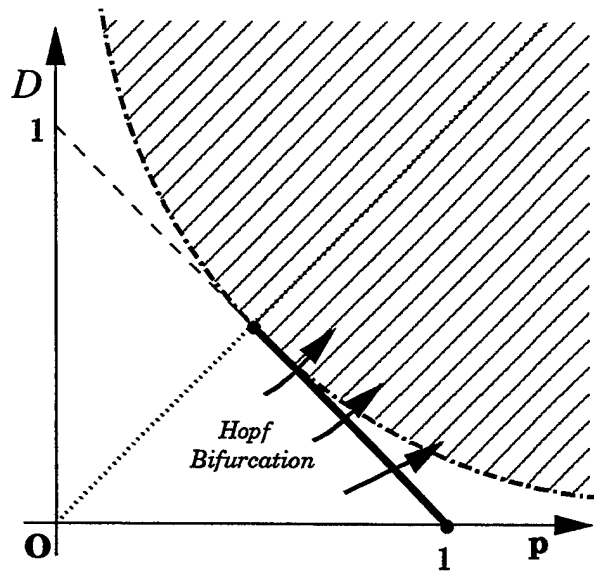
Figure 3: Parameter space.

which is equivalent to looking for the particular solution

$$\begin{pmatrix} e \\ f \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix}_0 + \epsilon \left[\begin{pmatrix} A(X,T) \\ B(X,T) \end{pmatrix} e^{i\omega t} + c.c. \right] + \epsilon^2 \begin{pmatrix} e \\ f \end{pmatrix}_2 + ..., \quad (13)$$

with the long α, X, and long time scale, T.

The linear evolution equation for the amplitude $\begin{pmatrix} A \\ B \end{pmatrix}$ is the Fourier-Laplace transform of the dispersion relation given by equation (9), (see Drazin & Johnson (1988), Fauve (1985) and Dodd et al. (1982)). By expanding equation (9) as a Taylor series this is easily obtained. The form of higher orders in the evolution equation is fixed by imposing translational invariance and considering solvability conditions after making the substitution equation (13). One of the first non-linear terms which are translationally invariant is $|A|^2B$ (and other similar terms) and the factor of proportionality is shown to be real by considering space reflective symmetry, (Fauve, 1985). The signs

of these parameters determine the criticality of the bifurcation (ie whether sub or super critical — if the bifurcation is subcritical then further terms should be sought). This equation is similar to the one dimensional complex Ginzberg-Landau equation

$$\partial_T A = \eta A + \nu \partial_X^2 A - \mu |A|^2 A. \tag{14}$$

Having derived the complex Ginzberg Landau equation, or similar in two variables, we can choose the parameters to obtain the behaviour that we require. Figure (4) shows the homoclinic orbit in the conceptual homogeneous phase space of e and f associated with a pulse, which is a consequence of a limit cycle associated with the Hopf bifurcation colliding with an external equilibrium point. This is described by the complex Ginzberg-Landau equation.
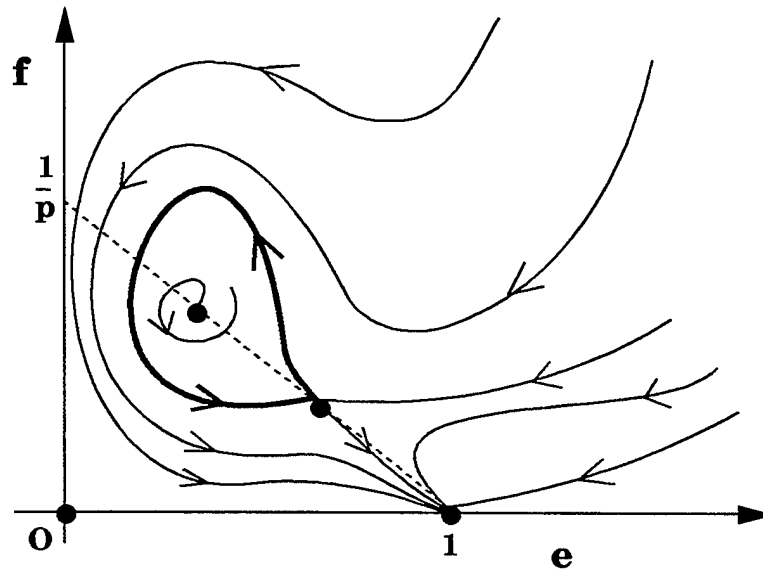
Figure 4: Conceptual phase portrait (2-D).

Bretherton & Spiegel (1983) describe a "soliton cascade", from the complex Ginzberg-Landau equation, which looks dramatically like the speciation of

the punctuated equilibria theory (figure 5). Pulses oscillate and suddenly split into two. Chaté reviews the complex Ginzberg-Landau equation and its parameter space in his 1993 paper.

There is, however, a problem. The pulses exist on a non-zero background population. That is to say that there is a non-zero number of species of every conceivable characteristic (e.g. giraffes with necks of infinite length). This is, perhaps, not very "realistic". Indeed, these solutions only exist when f is non-zero at plus and minus infinity, and as this is not realistic we can rule out these undesirable solutions. It turns out that the kernels in the original equation have a far greater effect than anticipated. They enable trajectories to escape from the otherwise stable mainfold at $(e, f) = (1, 0)$ and perform homoclinic orbits (see later).

4 Simulation

Since the kernel distributions are arbitrary, I can choose them to be defined on a closed interval as a truncated normal distribution and zero elsewhere. Characteristic space is discretized and the time derivative is approximated using a Runge-Kutta scheme. (To check the stability of the numerical method I also used a crude approximation for the time derivative and obtained very similar results. Overall, the behaviour of the system was found to be insensitive to the numerical scheme used). I then get a finite, implicit set of algebraic equations to be calculated at each time step.

$$\begin{pmatrix} \tilde{e}_i \\ \tilde{f}_i \end{pmatrix}(t) = \begin{pmatrix} e_i \\ f_i \end{pmatrix}(t) + \mathbf{G}(e_{i-\sigma}, ..., e_{i+\sigma}, f_{i-\sigma}, ..., f_{i+\sigma}; t)\delta t, \quad (15)$$

$$\begin{pmatrix} e_i \\ f_i \end{pmatrix}(t + \delta t) = \begin{pmatrix} e_i \\ f_i \end{pmatrix}(t) + (\mathbf{G}(e_{i-\sigma}, ..., e_{i+\sigma}, f_{i-\sigma}, ..., f_{i+\sigma}; t)$$
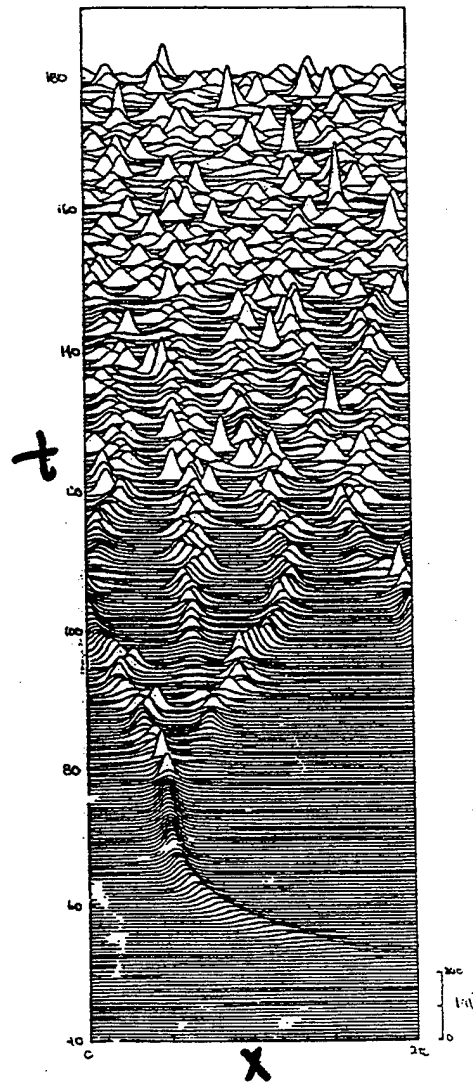
Figure 5: Soliton cascade, from Bretherton & Spiegel (1983).

$$+ \quad G(\tilde{e}_{i-\sigma}, ..., \tilde{e}_{i+\sigma}, \tilde{f}_{i-\sigma}, ..., \tilde{f}_{i+\sigma}; t)\Big) \frac{\delta t}{2}, \tag{16}$$

for some function G and some range of the kernels σ. I choose σ to be twice the standard deviation of the normal distributions. Hence, for equal distributions for the three kernels, the computer time is proportional to $n\sigma^3$, where n is the number of points in the simulation. The initial conditions determine the solution for all time. That is to say, by guessing an initial distribution of e and f, and if it persists and travels then I have found a travelling pulse. The boundary conditions are set up so as not to affect the region in characteristic space of interest. That is to say, they are set to $(e, f) = (1, 0)$ sufficiently far away from the pulse(s). In figure (6) the initial

conditions are shown. The environment is not set to zero in its "anti-pulse" but very close to zero, as its growth rate is proportional to itself. The species is set up as an exponential curve. Figure (7) shows a contour plot of e and
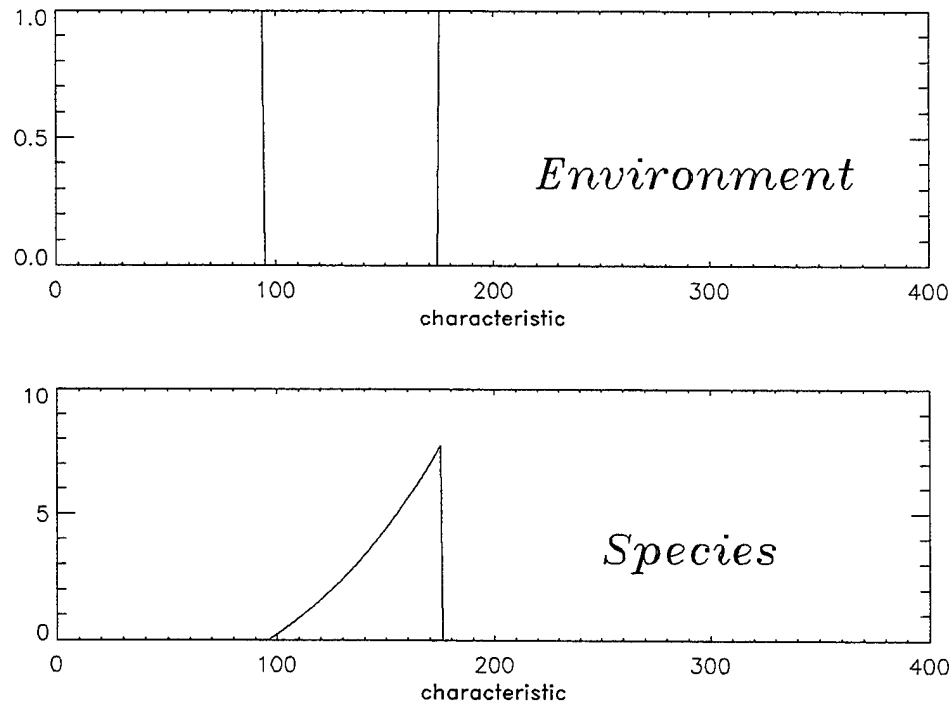
Figure 6: Initial conditions.

shows how this pulse persists for all time, with the species eating the fresh edge of the environment. Species that stay in the same place soon destroy the environment and become extinct. After some time the environment recovers back to full capacity. The speed of the pulse is governed by a balance between diffusive terms and the environmental holding capacity. By increasing $\mathcal{D}$ sufficiently the pulse does not persist, but decays to zero. Decreasing $\mathcal{D}$ or p can increase the pulse height. There are a number of ways to affect the

behaviour of the pulses. Increasing the variance of the kernel increases the
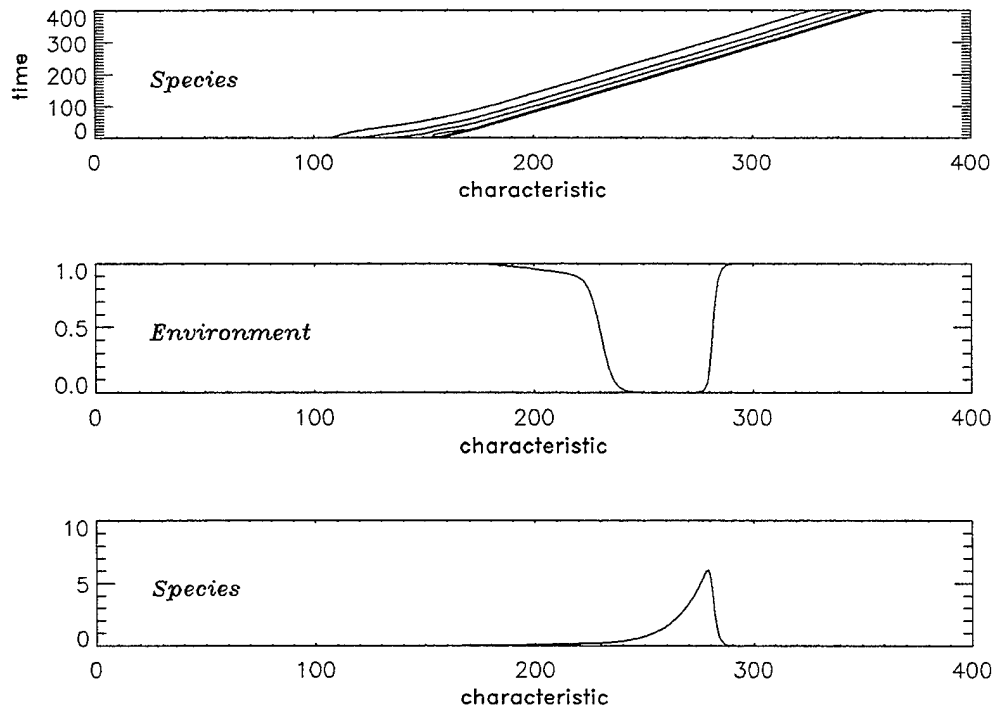
Figure 7: A pulse that persists.

speed of the pulses and makes them smoother, as does increasing the variance
of the birth kernel. However, increasing the variances of the relationship or
birth distributions, or decreasing $\mathcal{D}$ or the variance of the mutation, can
make the pulse split (see for example figure 8).

If enough species are left on or have diffused onto the tail end of the pulse
(even if they are dying out) and the environment starts to recover then they
can start eating the recovering environment. They will increase in number
if there are sufficiently many of them to have survived the hard times of low
environmental capacity but sufficiently few of them not to inhibit the envi-
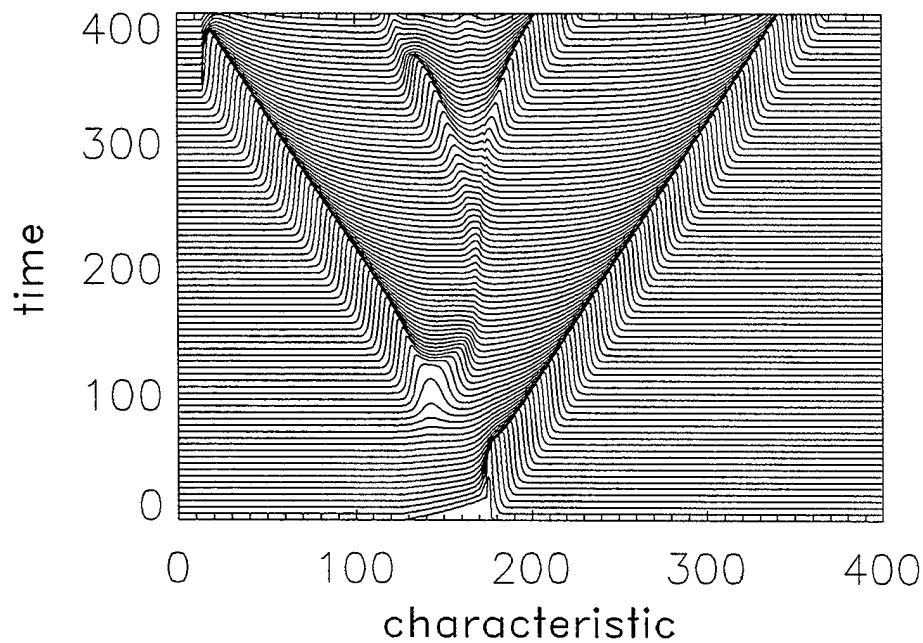
Figure 8: A splitting pulse.

ronmental recovery. Figure (9) shows an example of multiple splittings, or a "speciation tree". For parameter ranges that allow only very low level pulses (comparable with the homogeneous solutions) a solution can exist where two pulses go left and right but leave behind a homogeneous, non-zero solution as described in the previous section. In fact for these parameter ranges (such as for relatively large p), this appears to be the preferred solution. In this case the oscillatory equilibrium point has become stable and draws in the local trajectories in phase space.

Collisions of pulses are also of interest. These can occur when two groups of species both evolve into the same region of characteristic space, and can

result in total extinction or a "soliton" phase lag interaction. For this model, these collisions are unlikely to be realistic as we are allowing breeding between individuals of the same charateristic, regardless of their genetic past. The splitting, however, looks remarkably like Eldredge & Goulds' (1972) punctuated equilibria with a definite gap between the speciating groups. That is, it looks like a faster mechanism than natural selection is involved in the speciation. The speciation is actually caused by a dying population, left behind and isolated from the main bulk of a species, that suddenly finds a recovering environmental niche to use.
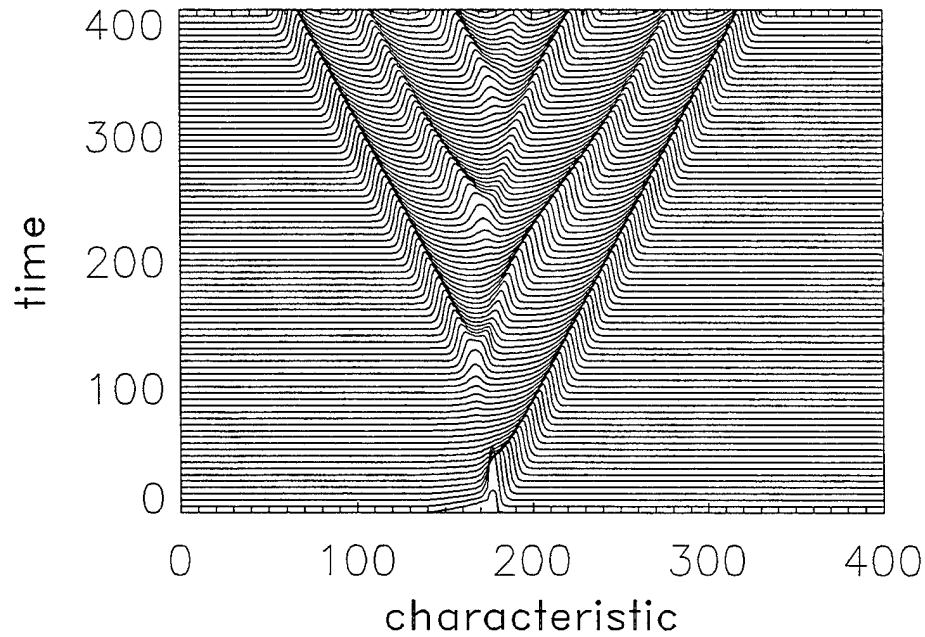
Figure 9: A pulse tree.

5 Dynamical system

Now reconsider the original integro-differential equations. If we assume the kernels are normal distributions then we can expand the integrals using the method of steepest descents (which boils down to Laplaces' method of expansions in Taylor series), to get

$$\partial_t e = e(1 - e) - pef, \tag{17}$$

$$\partial_t f = ef^2 + E^+ ef\partial_\alpha^2 f + E^- e(\partial_\alpha f)^2 - \mathcal{D}f + m_2\partial_\alpha^2 f + h.o.t., \tag{18}$$

where E^+, E^- and m_2 are constants determined by the moments of the kernels. That is

$$m_n = \int \frac{m(z)z^n}{n!} dz, \tag{19}$$

$$E^+ = 2B_2 + \frac{R_2}{2} \tag{20}$$

and

$$E^- = 2B_2 - \frac{R_2}{2}, \tag{21}$$

where R_n and B_n are defined in a similar way to m_n.

Assuming adequate exponential drop off we can approximate the infinite dimensional system by truncating the expansions. Next, we look for a particular type of solution; a travelling wave solution of the form

$$\begin{pmatrix} e \\ f \end{pmatrix}(\alpha, t) = \begin{pmatrix} e \\ f \end{pmatrix}(\alpha + ct), \tag{22}$$

where c is the pulse speed and e and f are allowed to vary also on a slow time scale. If this time scale becomes too fast then the travelling wave approximation breaks down.

Hence, I get the finite dimensional dynamical system

$$e' = \frac{e}{c}(1 - e) - \frac{p}{c}fe, \tag{23}$$

$$f' = g, \tag{24}$$

$$g' = \frac{\mathcal{D}f + cg - E^- eg^2 - ef^2}{E^+ ef + m_2}. \tag{25}$$

The phase portrait for this system is similar to that of the 2-d system (figure 10) except that it has an extra dimension, due to the diffusive properties of the system, which allows trajectories to escape from the once stable manifolds. For the system before truncation, the diffusive properties of the equations potentially lead to an infinite number of dimensions in which the trajectories can escape from the stable manifolds. For a set of parameter
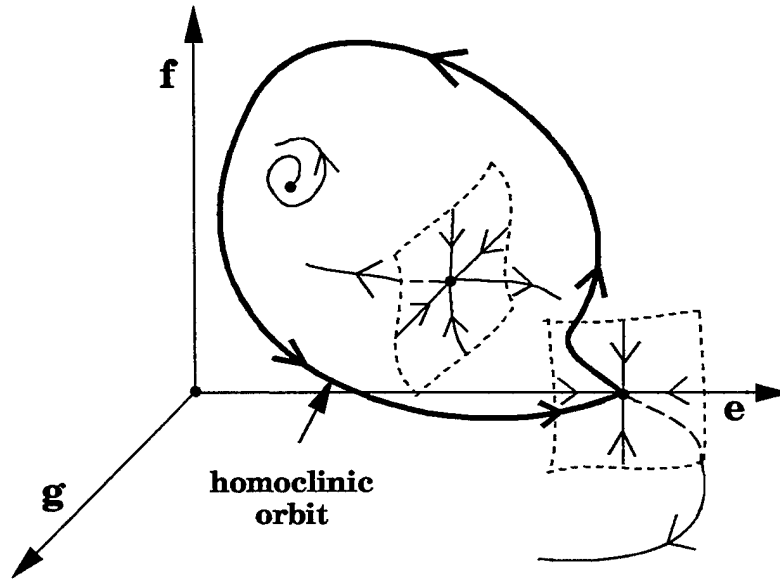
Figure 10: Phase portrait - homoclinic orbit (3-d).

values there exists a homoclinic orbit from and to the equilibrium point at $(e, f, g) = (1, 0, 0)$ as long as the trajectory avoids the stable manifold of a hyperbolic point, as seem in figure (10). Alternatively, if the trajectory ends up on the wrong side of the hyperbolic point then it can get pushed away from $(1, 0, 0)$ and a hump can form on the tail end of the pulse, and when

this becomes significant, breaks down the travelling wave assumption. If this occurs then there is a transition period before two travelling pulse solutions are generated, one with speed c and another with speed $-c$ (which can be considered as a pulse with speed c under the symmetry $x + ct \longrightarrow -(x + ct)$). This leads to two homoclinic orbits along the same trajectory, but in opposite directions. This explanation is backed up by plotting the numerical results on the 3-d phase portrait (figures 11 and 12).
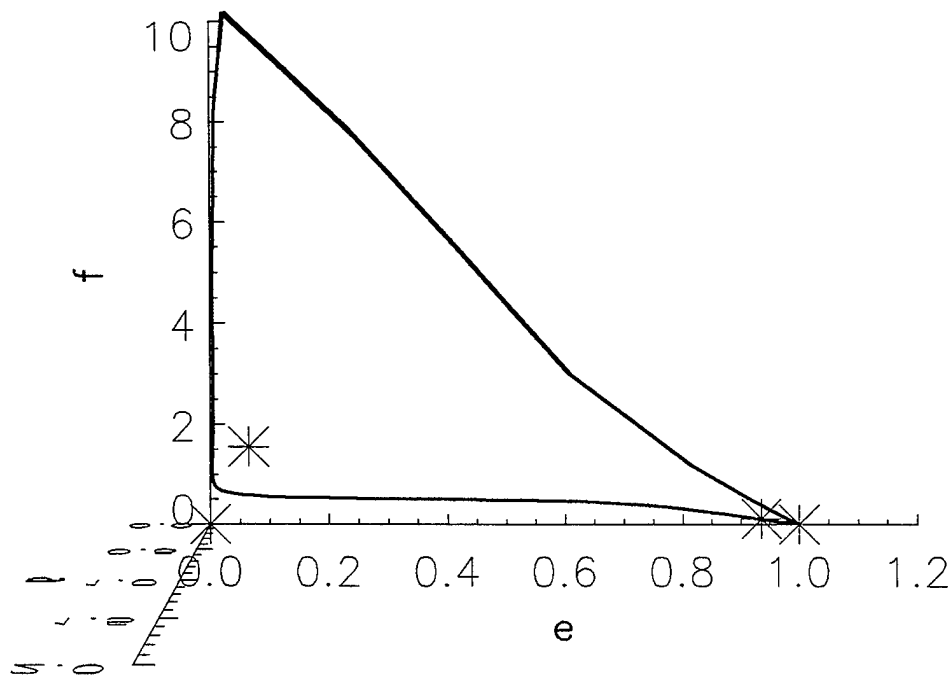
Figure 11: Numerical phase portrait - homoclinic orbit. Thicker lines are in the foreground.
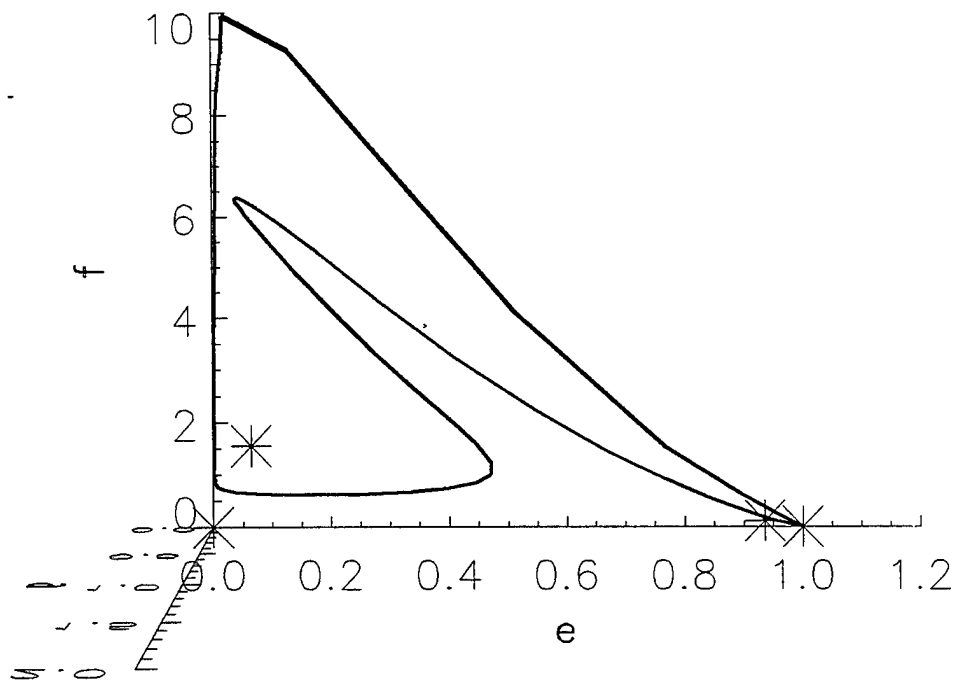
Figure 12: Numerical phase portrait - the moment of splitting. Thicker lines are in the foreground.

6 Discussion

I have demonstrated that for a relatively simple model, where the species needs and affects a particular part of the environment and with a sufficient spread of mutation, even if there are very low levels of beneficial mutation, I can get periods of gradual evolution followed by abrupt speciations. In fact, these speciations can be triggered by changing the parameters or by the manipulation of environmental space, but more importantly, often happen by themselves at regular stages in a pulses' life. It is also possible, for the correct parameter ranges, to get a more gradual speciation, but this often

only happens for symmetric initial conditions. In general, the speciation is reminiscient to both the punctuated equilibria of Eldredge & Gould (1972) and aspects of gradualism.

Obviously, a species would normally have many distinguishing characteristics and may abuse the environment in many different ways. Therefore, an extension to the present work would be to consider a higher dimensional characteristic space, and possibly to include the effects of geographical isolation.

A procedure for analysing the integro-differential equations as a dynamical system is presented in this paper which could conceivably be used for other similar systems. For our model, the governing equations are of interest in their own right, regardless of how they represent reality, and under a fuller analysis, could produce even more interesting behaviour.

References

[1] Bretherton, C S & Spiegel, E A. 1983. Intermittency through modulational instability. Physics Letters. **96A**:152 - 156.

[2] Chaté, H. 1993. Spaciotemporal patterns in nonequilibrium complex systems. Santa Fe Institute Series in the Sciences of Complexity, Addison Isley.

[3] Dawkins, R. 1989. The selfish gene. New edition. Oxford University Press.

[4] Dodd, R K, Eilbeck, J C, Gibbon J D & Morris H C. 1982. Solitons and non-linear wave equations. Academic Press, London.

[5] Drazin, P G & Johnson, R S. 1988. Solitons: an introduction. Cambridge university press.

[6] Eldredge & Gould. 1972. Punctuated Equilibria : An alternative to phyletic gradualism. Models in Paleobiology. (T.J.M Schopf, ed.) pp82 - 115. Freeman, Cooper, San Francisco.

[7] Ewens, W J. 1969. Population Genetics. Methuen, London.

[8] Fauve. 1985. Large scale instabilities of cellular flows. WHOI technical report, Woods Hole, Massachussetts.

[9] Futuyma, D J. Evolutionary Biology. 2nd ed. Sinauer Associates, Inc. Sunderland, Massachussetts.

[10] Hecht, Eldredge & Gould. 1972. Morphological transformation, the fossil record, and the mechanisms of evolution : A debate. Evolutionary Biology. Vol 7:8.

[11] Levin, S A & Segel, L A. 1982. Models of the influence of predation on aspect diversity in prey populations. Journal of Mathematical Biology. 14:253 - 284.

[12] O'Brien, P. 1985. A genetic model with mutation and selection. Mathematical Biosciences. 73: 239 - 251.

[13] Rensberger, B. 1981. The evolution of evolution. Mosaic. September/October.

[14] Roughgarden, J. 1972. Evolution of niche width. The American Naturalist. Vol. 106, No. 952.

[15] Slatkin, M & Lande, R. 1976. Niche width in a fluctuating environment-density independent model. The American Naturalist. Vol. 110, No. 971.

[16] Spencer, H & Barakat, R. 1992. Random genetic drift and selection in a triallelic locus: a continuous diffusion model. Mathematical Biosciences. **108**:127 - 139.

[17] Veronka, R & Keller, J B. 1975. Asymptotic analysis of stochastic models in population genetics. Mathematical Biosciences. **25**:331 - 362.

[18] Weiner. 1994. The Beak of the Finch. Knopf, NewYork/ Jonathan Cape, London. Pp. 332.

Pattern formation in slime mold aggregation
Amar Gandhi

1   Introduction

In response to stress, *Dictyostelium discoideum* amoebae signal to each other chemically using cyclic adenosine 3′, 5′-monophosphate (cAMP). In a field of signaling amoebae spread over a surface, pulses of extracellular cAMP travel across the field in form of either expanding concentric circular waves or rotating spiral waves (see figure 1). As the waves pass periodically through the field of independent amoebal cells, they stimulate a chemotactic movement of the amoebae towards the center of the pattern (the origin of the circles or the pivot point of the spiral). Eventually all the amoebae within the domain of a single pattern aggregate at the center to form a multicellular slug, which goes on to form a fruiting body. The aggregation phase of *Dictyostelium discoideum* is often taken as a paradigm for developmental biology for we see the development of a spatiotemporal periodic chemical prepattern that induces a pattern of morphogenetic movements culminating in the formation of an organized multicellular tissue.
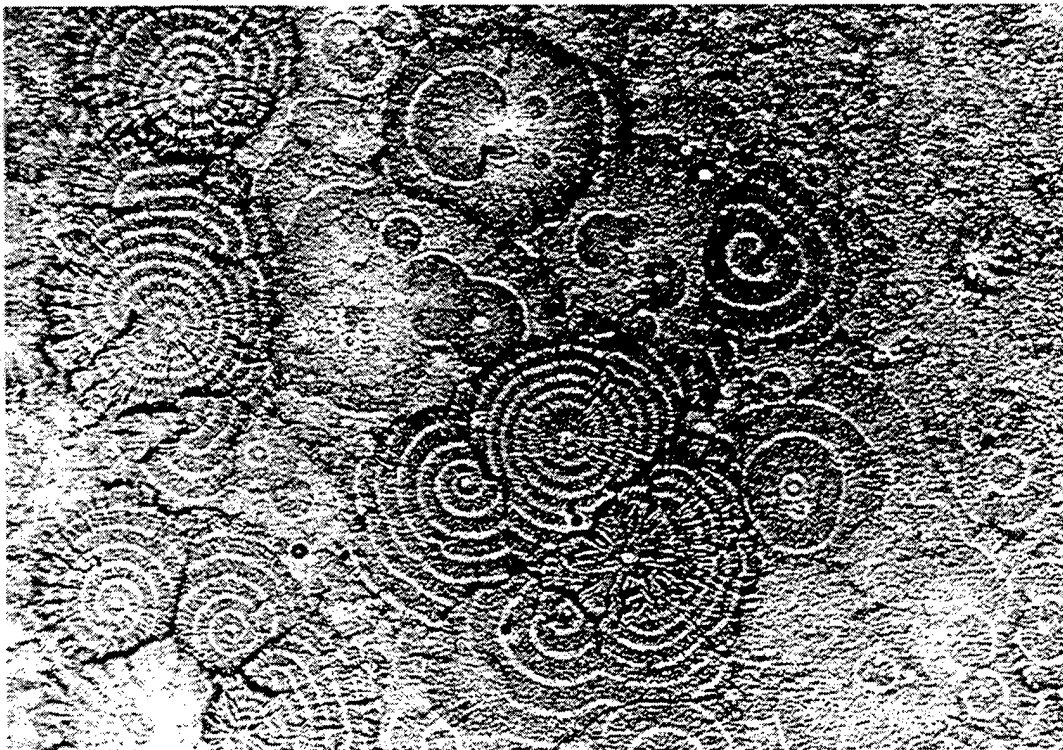
Figure 1: Target and spiral patterns in *Dictyostelium discoideum*. (reprinted from Newell, P. C. (1983) by courtesy of Marcel Dekker Inc.)

In this paper, we study a model based on molecular mechanism of the cAMP relay response that describes the observed spatiotemporal chemical patterns. The model neglects any chemotactic motion of amoebae and so is only valid in the initial stages of aggregation when there is still a uniform distribution of amoebae. We proceed as follows. In section 2, we discuss the reaction-diffusion equation used to model the cAMP waves. In section 3, we study the local kinetics by setting the diffusivity to zero. This allows us to understand the dynamics occurring locally at a point in absence of coupling from neighboring points. In section 4, we include the effect of diffusion and derive the dispersion relation. In section 5, we derive an amplitude equation to serve as a model for behavior near bifurcation. In sections 6 through 8, we take a more heuristic approach and study how the local kinetics are affected by coupling with other points through diffusion to give rise to plane waves, target patterns and spiral waves respectively. Section 9 presents evidence for subcriticality with respect to formation of spiral waves.

2   Reaction-diffusion equation

2.1   cAMP chemistry

When *Dictyostelium discoideum* amoebae are left to starve on an agar surface, they communicate with each other using a chemical messenger, cAMP. Individual cells receive the signal by binding extracellular cAMP to a membrane receptor, and this binding stimulates the synthesis of cAMP within the cell. Newly synthesized cAMP is transported to the extracellular medium and in this fashion the chemical signal is amplified. Amplification of the signal is limited by the fact that, on prolonged exposure to cAMP, the membrane receptor becomes desensitized. In the absence of cAMP synthesis, the concentration of cAMP (both inside and outside the cell) decreases by the action of an hydrolyzing enzyme. According to this picture of cAMP chemistry, confirmed by Martiel & Goldbeter (1987), the relevant variables to explain the spatiotemporal patterns are the fraction of membrane receptors in active form (ρ), the extracellular cAMP concentration (γ), and the intracellular cAMP concentration (β) which vary as follows :

$$
\begin{array}{ccccc}
\text{Rate of change of} & & & \text{desensitization} & \text{resensitization} \\
\text{active form of receptor} & = & - & \text{of receptor} & + \;\text{of receptor} \\
\partial\rho/\partial t & = & - & f_1(\gamma)\rho & + \; f_2(\gamma)(1-\rho) \qquad (1)
\end{array}
$$

$$
\begin{array}{ccccc}
\text{Rate of change of} & = & \text{secretion} & \text{extracellular} & \text{random} \\
\text{extracellular cAMP} & = & \text{by the cells} & -\;\text{hydrolysis} & + \;\text{diffusion} \\
\partial\gamma/\partial t & = & (k_t/h)\beta & -\;\; k_e\gamma & + \quad D\Delta\gamma \qquad (2)
\end{array}
$$

$$\begin{array}{llll}
\text{Rate of change of} & = & \text{synthesis} & - \text{secretion} & + \text{intracellular} \\
\text{intracellular cAMP} & & \text{in the cells} & \text{by the cells} & \text{hydrolysis}
\end{array}$$

$$\partial\beta/\partial t \quad = \quad q\Phi(\rho,\gamma) \quad - \quad k_t\beta \quad - \quad k_i\beta \qquad (3)$$

D is the diffusion coefficient of extracellular cAMP. The constants k_t, k_e, q and k_i are rate constants associated with the metabolism of cAMP, and h is the ratio of extracellular volume to intracellular volume. The rate functions

$$f_1(\gamma) = \frac{k_1 + k_2\gamma}{1+\gamma}, \quad f_2(\gamma) = \frac{k_{-1} + k_{-2}c\gamma}{1 + c\gamma}$$

describe the receptor kinetics and

$$\Phi(\rho,\gamma) = \frac{\lambda_1 + Y^2}{\lambda_2 + Y^2}, \quad Y = \frac{\rho\gamma}{1+\gamma}$$

describes the cAMP synthesis in cells where the k's, λ's and c are constant parameters.

2.2 Reduction to (ρ,γ) system

Following Martiel & Goldbeter (1987) and Tyson et. al. (1989), we rewrite eqns. (1),(2) and (3) as :

$$\frac{\partial\rho}{\partial t} = -f_1(\gamma)\rho + f_2(\gamma)(1-\rho) \qquad (4)$$

$$\frac{1}{k_e}\frac{\partial\gamma}{\partial t} = \frac{k_t}{k_e h}\beta - \gamma + \frac{D}{k_e}\Delta\gamma \qquad (5)$$

$$\frac{1}{k_i + k_t}\frac{\partial\beta}{\partial t} = \frac{q}{k_i + k_t}\Phi(\rho,\gamma) - \beta \qquad (6)$$

Assuming $1/(k_i + k_t) \ll 1/k_e < 1$ i.e., β reacts much faster than ρ and γ, eqn. (6) becomes

$$\beta = \frac{q}{k_i + k_t}\Phi(\rho,\gamma) \qquad (7)$$

This allows us to eliminate β from the model yielding

$$\frac{\partial\rho}{\partial t} = -f_1(\gamma)\rho + f_2(\gamma)(1-\rho) \qquad (8)$$

$$\frac{\partial\gamma}{\partial t} = \frac{k_1}{e}[s\Phi(\rho,\gamma) - \gamma] + D\Delta\gamma \qquad (9)$$

where $e = k_1/k_e$ and $s = (qk_t)/(hk_e(k_i + k_t))$. In this 2-variable model, the parameters are k_1, k_2, k_{-1}, k_{-2}, c, λ_1, λ_2, e, s, and D. We will use this model for the rest of this paper with the following parameters fixed :

$$
\begin{aligned}
k_1 &= 0.036, & k_2 &= 0.666, \\
k_{-1} &= 0.36, & k_{-2} &= 0.0033, & c &= 10 \\
\lambda_1 &= 0.001, & \lambda_2 &= 2.4
\end{aligned}
$$

The free parameters are e, s and D.

3 Reaction kinetics alone — No diffusion

In this section, we study how the local kinetics depend on the parameters. To do this, we set $D = 0$ which allows us to study a single point in space decoupled from all others. Since we are studying the behavior at a single point, ρ and γ will only depend on time. Once we have understood the behavior of the system locally at a point, we will turn diffusion on in the following sections and see the effect of coupling the points together.

The fixed point (ρ_0, γ_0) is at the intersection of the nullclines

$$
\begin{aligned}
\rho\text{--nullcline} &= \left\{ (\rho, \gamma) : \frac{d\rho}{dt} = 0, \rho \geq 0, \gamma \geq 0 \right\} \\
\gamma\text{--nullcline} &= \left\{ (\rho, \gamma) : \frac{d\gamma}{dt} = 0, \rho \geq 0, \gamma \geq 0 \right\}
\end{aligned}
$$

Figure 2 shows the nullclines, the fixed point and some sample trajectories for $s = 49, e = 0.01$. For this set of parameters, the system shows excitable behavior — when the system is perturbed off the fixed point, it undergoes a wild excursion before returning to it.

Figure 3 shows the nullclines, the fixed point and some sample trajectories for $s = 50, e = 0.01$. For this set of parameters, the system has a limit cycle. The fixed point from the previous figure has lost its stability to the limit cycle through a hopf bifurcation.

To understand behavior around the fixed point (ρ_0, γ_0), we let

$$
\rho(t) = \rho_0 + \rho'(t), \quad \gamma(t) = \gamma_0 + \gamma'(t)
$$

and substitute into equations (8) and (9). Retaining only the linear terms, we obtain

$$
\begin{bmatrix} \rho_t' \\ \gamma_t' \end{bmatrix} = \mathbf{A} \begin{bmatrix} \rho' \\ \gamma' \end{bmatrix} \tag{10}
$$

Eigenvalues of $\mathbf{A}$ will tell us about the stability properties of the fixed point.
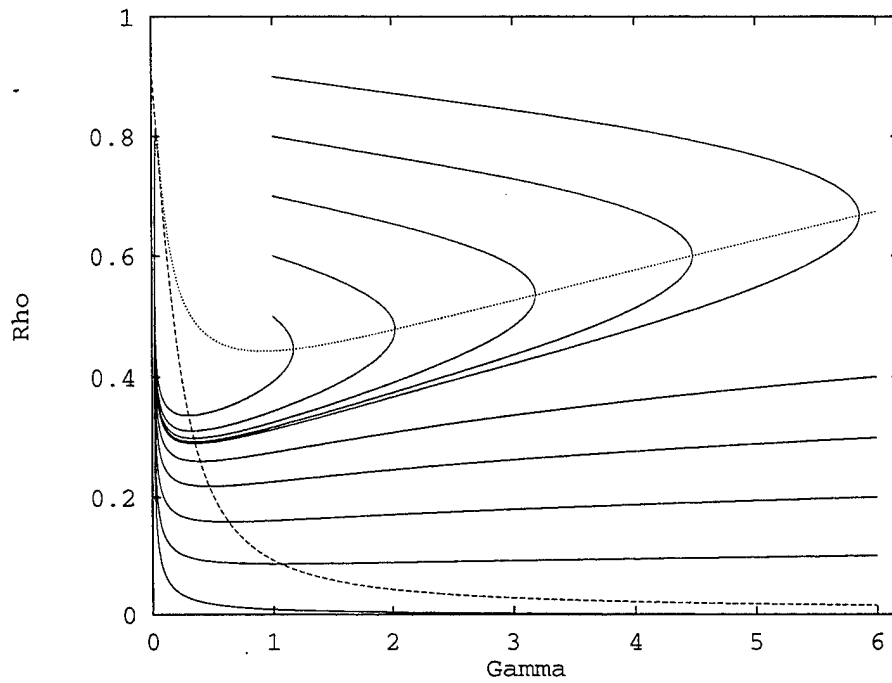
Figure 2: System displays excitable behavior for $s = 49, e = 0.01$. γ-nullcline is the dotted curve and the ρ-nullcline is the dashed curve.
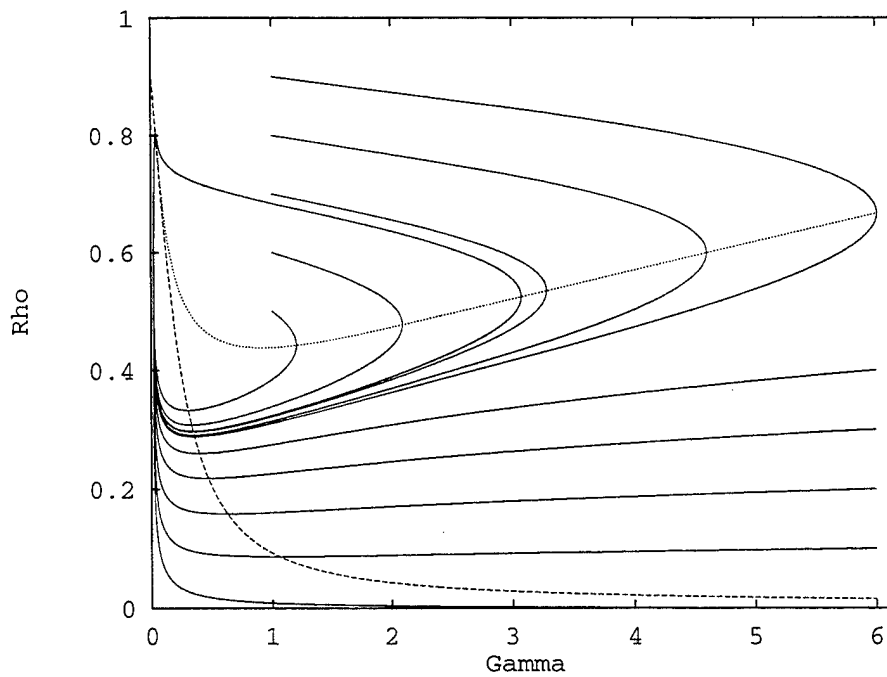
Figure 3: System has a limit cycle for $s = 49, e = 0.01$. γ-nullcline is the dotted curve and the ρ-nullcline is the dashed curve.

3.1 Effect of varying s

Figure 4 shows the locus of the fixed point while figure 5 shows the locus of the eigenvalues of linearization around the fixed point as s is increased from 10 to 300. Figure 6 shows the real part of the eigenvalues as a function of s. The eigenvalues start out with negative real parts and zero imaginary parts. As s is increased, they move towards each other. Upon collision at $s \sim 45$, they acquire imaginary parts and move towards the positive real half-plane. The hopf bifurcation occurs at $s \sim 50$. Soon after this, they collide again and become all real. They begin to move away from each other but then stop and start to move towards each other again. At $s \sim 60$, they collide once more and acquire imaginary parts. They move towards the negative real half-plane and another hopf bifurcation takes place around $s \sim 220$. e is taken to be 0.01.
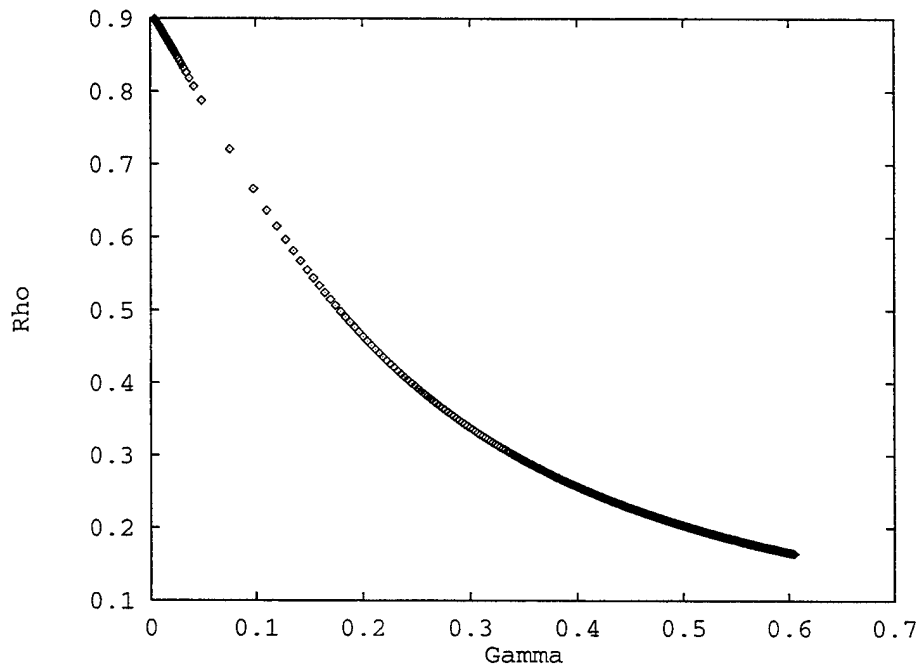
Figure 4: Locus of the fixed point as s is increased from 10 to 300 in increments of 1. The fixed point corresponding to $s = 10$ is in the upper left-hand corner and it moves towards the lower right-hand corner as s is increased. $e = 0.01$.
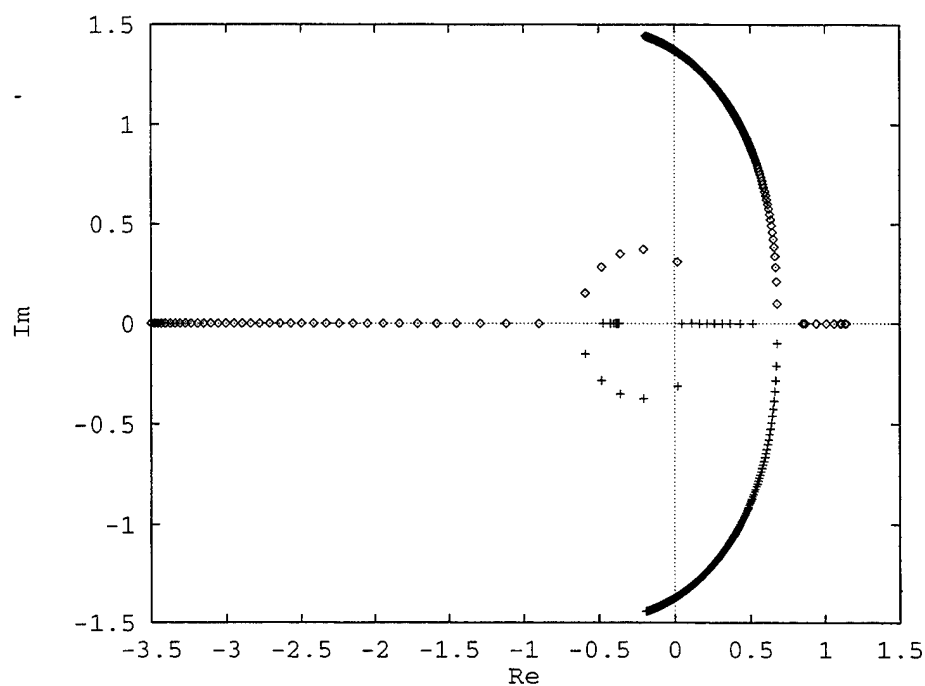
Figure 5: Locus of the eigenvalues of linearization around the fixed point as s is increased from 10 to 300 in increments of 1. $e = 0.01$.
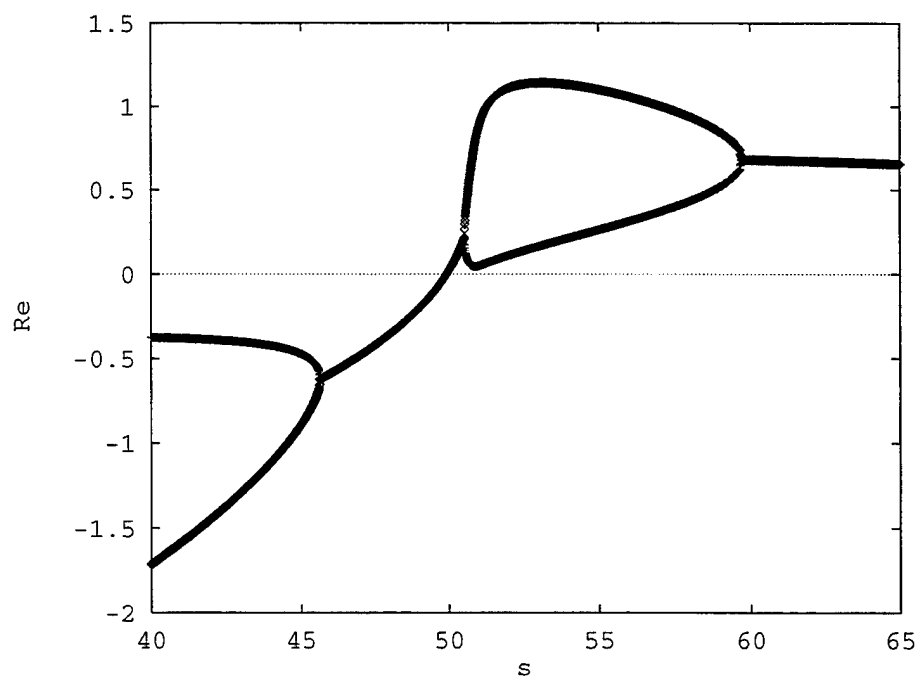
Figure 6: Real part of the eigenvalues of linearization around the fixed point as s is increased from 45 to 60 in increments of 0.01. $e = 0.01$.

3.2 Effect of varying e

Notice that e is the only parameter that can be varied without affecting the location of the fixed point. It does affect the stability of the fixed point, however. Figure 7 shows the dependence of the eigenvalues of linearization around the fixed point upon varying e for $s = 50$. We observe that the linear growth rate is zero for $e = e_c \equiv 0.01129$.
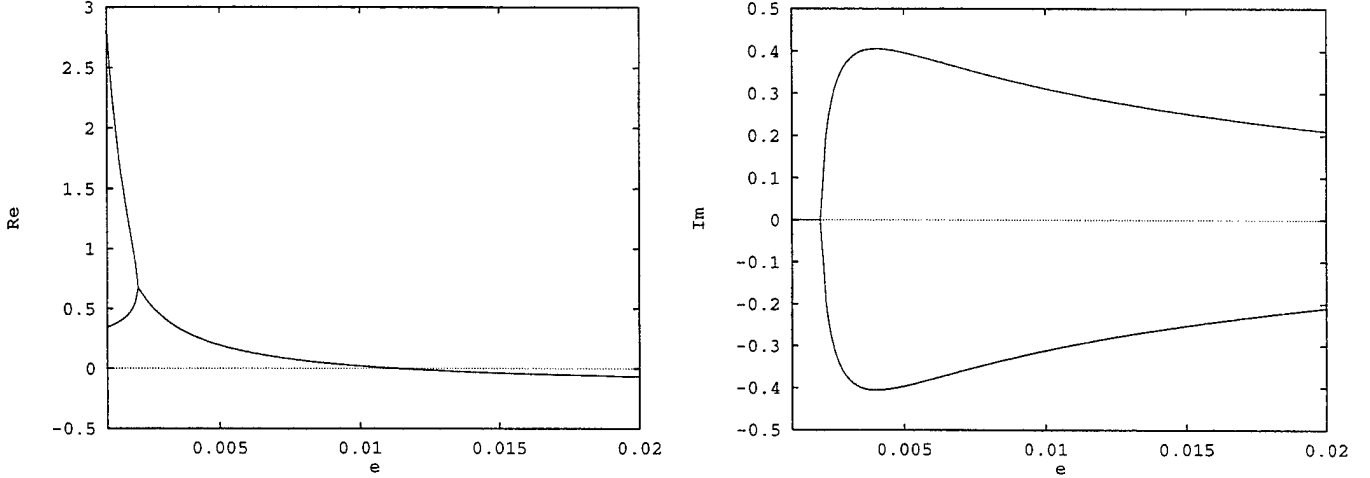
Figure 7: Dependence of the eigenvalues of linearization around the fixed point upon varying e for $s = 50$.

4 Dispersion relation

In the previous section we saw that in the absence of diffusion or equivalently for the uniform state, the linear growth rate vanishes for $e = e_c, s = 50$. With these parameters, we turn on diffusion (by setting $D = 0.0001$) and perform the linear stability analysis to determine how the linear growth rate varies for disturbances with wavenumber k. We let

$$\rho(\mathbf{x}, t) = \rho_0 + \rho' \exp[i\mathbf{k} \cdot \mathbf{x} + \sigma t], \quad \gamma(\mathbf{x}, t) = \gamma_0 + \gamma' \exp[i\mathbf{k} \cdot \mathbf{x} + \sigma t], \quad |\mathbf{k}| = k$$

and substitute into equations (8) and (9). Retaining only the linear terms, we obtain

$$\begin{bmatrix} \sigma \rho' \\ \sigma \gamma' \end{bmatrix} = \mathbf{A} \begin{bmatrix} \rho' \\ \gamma' \end{bmatrix} + \begin{bmatrix} 0 \\ -Dk^2 \gamma' \end{bmatrix}$$

or

$$\mathbf{B} \begin{bmatrix} \rho' \\ \gamma' \end{bmatrix} = 0 \quad \text{where} \quad \mathbf{B} = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma + Dk^2 \end{bmatrix} - \mathbf{A} \tag{11}$$

Setting the determinant of $\mathbf{B}$ to zero will yield the growth rate σ. Figure 8 shows the dependence of the linear growth rate upon k. We see that the most unstable mode corresponds to $k = 0$ and the instability leads to uniform oscillations.

295
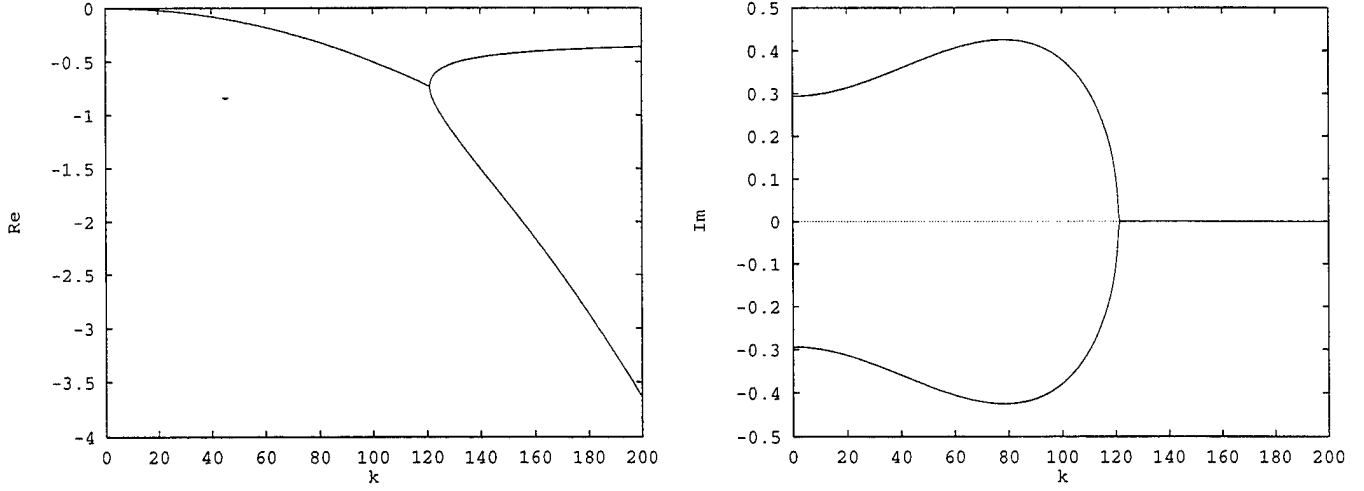
Figure 8: Dispersion relation, $\sigma(k)$ for $e = e_c, s = 50, D = 0.0001$.

5   Complex Ginzburg-Landau Equation

In this section, we study the behavior close to bifurcation by the amplitude equation. This is done using the multiple scale approach.

5.1   Control parameter and scaling

Figure 9 shows the change in dispersion relation as e is perturbed off of e_c. As e is increased, both $\sigma_{\text{real}}(k = 0)$ and $\sigma_{\text{imag}}(k = 0)$ decrease.

We define the control parameter

$$\delta = e_c - e.$$

For positive values of δ, $\sigma_{\text{real}}(k)$ curve is lifted up introducing a small band of unstable wavenumbers near zero. $\sigma_{\text{real}}(k = 0)$ is proportional to δ and since $\left.\frac{\partial \sigma(k)}{\partial k}\right|_{k=0} = 0$, the number of unstable modes will be proportional to $\sqrt{\delta}$. For the perturbative expansion, we define

$$\delta = \text{sgn}(\epsilon)|\epsilon|^2$$

where ϵ is the expansion parameter. Near the bifurcation, i.e., for $|\epsilon| \ll 1$, we wish to separate fast and slow scales for x and t. We therefore define

$$X = |\epsilon|x, \quad Y = |\epsilon|y, \quad T = |\epsilon|^2 t$$
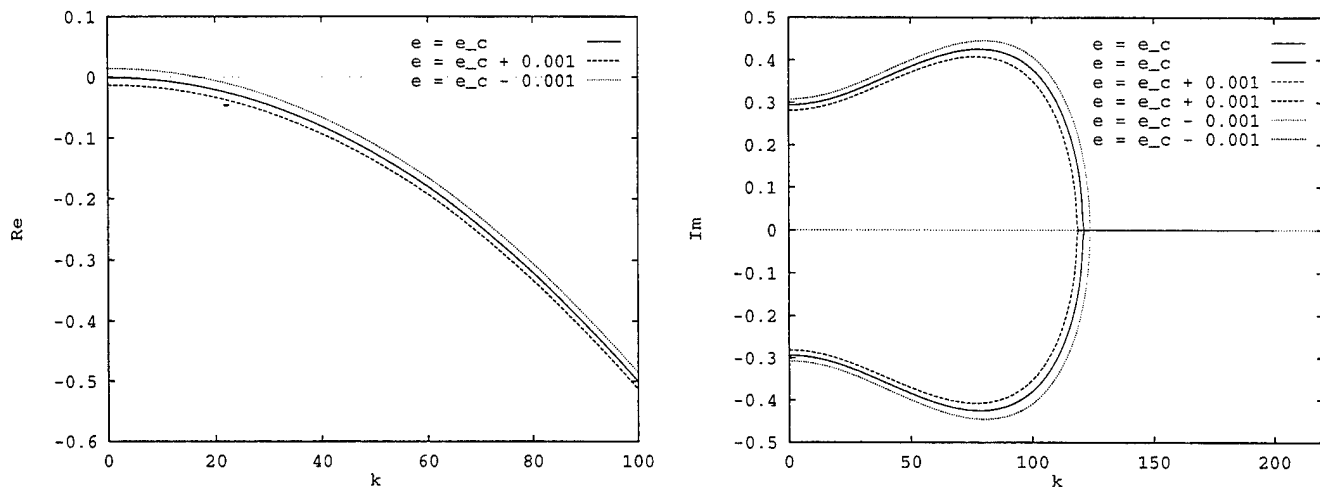
Figure 9: Dispersion relation, $\sigma(k)$ for $s = 50, D = 0.0001, e = e_c$ (solid curve), $e = e_c + 0.001$ (dashed curve), $e = e_c - 0.001$ (dotted curve).

5.2 Preliminaries

We write equations (8) and (9) formally as

$$\partial_t \mathbf{u} = \mathbf{F}(\mathbf{u}) \quad \text{using} \quad \mathbf{u} = \begin{bmatrix} \rho \\ \gamma \end{bmatrix}$$

At the fixed point,

$$\mathbf{u} = \mathbf{u_0} = \begin{bmatrix} \rho_0 \\ \gamma_0 \end{bmatrix}, \quad \partial_t \mathbf{u_0} = \mathbf{F}(\mathbf{u_0}) = \mathbf{0}.$$

We are interested in studying the behavior near bifurcation and so we pick a fixed point for which $\delta = 0$. Let $\mathbf{M}$ represent the linearization of $\mathbf{F}(\mathbf{u})$ about such a fixed point i.e.,

$$\mathbf{M} = \left. \frac{d\mathbf{F}(\mathbf{u})}{d\mathbf{u}} \right|_{\mathbf{u}=\mathbf{u_0}}$$

$\mathbf{M}$ has eigenvalues $(i\omega_0)$, $(i\omega_0)^*$ and eigenvectors $\mathbf{E}, \mathbf{E}^*$.

The solution of $\partial_t \mathbf{u} = \mathbf{M}\mathbf{u}$ is $\mathbf{u} = A \exp(i\omega_0 t)\mathbf{E} + \text{c.c.}$. Hence, $\mathbf{L} = (\partial_t - \mathbf{M})$ has eigenvectors $\mathbf{E}, \mathbf{E}^*$ with *zero* eigenvalues.

5.3 ϵ-expansion

We expand

$$\mathbf{u} = \mathbf{u_0} + |\epsilon|\mathbf{u_1} + |\epsilon|^2\mathbf{u_2} + \cdots$$

297

and collect orders of $|\epsilon|$:

$O(1)$

$$\mathbf{F}(\mathbf{u_0}) = 0$$

which is the fixed point condition.

$O(|\epsilon|)$

$$\mathbf{L}\mathbf{u_1} = 0$$

has the solution

$$\mathbf{u_1} = A(X, Y, T)\exp(i\omega_0 t)\mathbf{E} + \text{c.c.}$$

$O(|\epsilon|^2)$

$$\mathbf{L}\mathbf{u_2} = [\cdot\cdot]A^2\exp(2i\omega_0 t) + \text{c.c.} + [\cdot\cdot]AA^*$$

has the solution

$$
\begin{aligned}
\mathbf{u_2} = \ & B(X, Y, T)\exp(i\omega_0 t)\mathbf{E} + \text{c.c.} \\
& [\cdot\cdot]A^2\exp(2i\omega_0 t) + \text{c.c.} \\
& [\cdot\cdot]AA^*
\end{aligned}
$$

$O(|\epsilon|^3)$

$$
\begin{aligned}
\mathbf{L}\mathbf{u_3} = \ & \mathbf{H}\exp(i\omega_0 t) + \text{c.c.} \\
& +[\cdot\cdot]\exp(2i\omega_0 t) + \text{c.c.} \\
& +[\cdot\cdot]\exp(3i\omega_0 t) + \text{c.c.} + [\cdot\cdot]
\end{aligned}
$$

This equation has a nontrivial rhs so the linear operator on the left must be inverted. Since this operator has vanishing eigenvalues we must impose a *solvability condition*, requiring that the vector on the right should not drive any eigenvector with zero eigenvalue. Fredholm theorem implies that

$$(\mathbf{H}, \mathbf{E}^{\dagger*}) = 0$$

where $\mathbf{E}^\dagger$, $\mathbf{E}^{\dagger*}$ are the eigenvectors with *zero* eigenvalues of the adjoint operator $\mathbf{L}^\dagger = -(\partial_t + \mathbf{M}^\dagger)$. This yields the complex Ginzburg-Landau equation

$$\frac{\partial A}{\partial T} = (\text{sgn}(\epsilon)\alpha + \beta D\Delta)A + \mu|A|^2 A \tag{12}$$

where

$$\alpha = -\left.\frac{\partial\sigma(k=0)}{\partial e}\right|_{e=e_c}, \quad \beta = -\frac{1}{2}\left.\frac{\partial^2\sigma}{\partial k^2}\right|_{k=0}$$

298

Recall that this expansion is done about a fixed point for which $\delta = 0$. The fixed point for $s = 50, e = e_c$ is one such point and the corresponding numerical values for the parameters in the complex Ginzburg-Landau equation are

$$
\begin{aligned}
\alpha &= 13.64 + i13.00 \\
\beta &= 0.50 - i0.53 \\
\mu &= 1235.93 - i2389.66
\end{aligned}
$$

5.4 Local behavior — No diffusion

Setting $D = 0$, we obtain for the local behavior

$$\frac{dA}{dT} = \text{sgn}(\epsilon)\alpha A + \mu|A|^2 A \tag{13}$$

The fixed points of $|A|$ as a function of $\text{sgn}(\epsilon)\text{Re}(\alpha)$ are shown in figure 10. We see that for negative values of $\text{sgn}(\epsilon)\text{Re}(\alpha)$, there is a stable fixed point at the origin and an unstable limit cycle. As $\text{sgn}(\epsilon)\text{Re}(\alpha)$ increases towards positive values, the limit cycle collapses onto the origin and the origin becomes unstable through a *subcritical* hopf bifurcation.
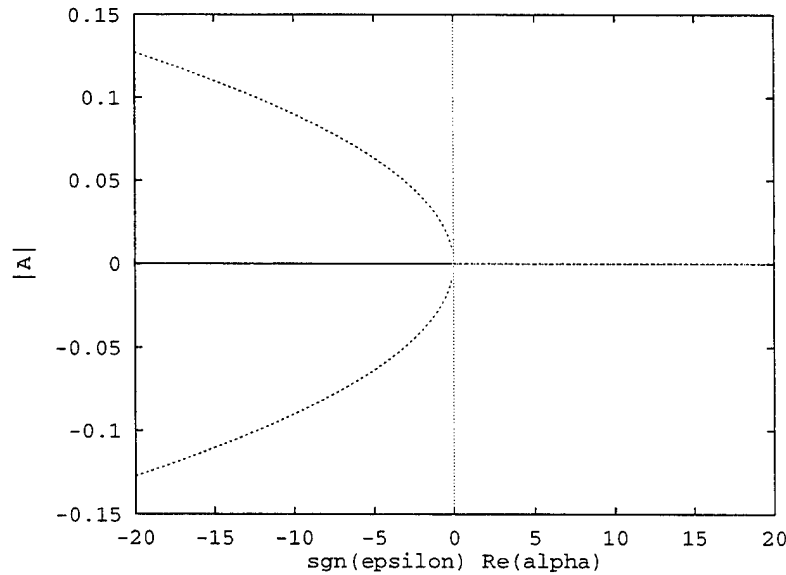
Figure 10: Fixed points of $|A|$ as a function of $\text{sgn}(\epsilon)\text{Re}(\alpha)$.

6 Plane waves

In this section, we begin our heuristic study of how the local kinetics (§3) are modified by coupling with other points through diffusion. Here, we study the formation and maintenance of one-dimensional plane waves. This understanding would be useful when studying circular and spiral waves in two dimensions because away from the center these waves have low curvature and resemble plane waves.

We would like to know if there exist any solutions to our system of the form

$$\rho = R(x + vt)$$
$$\gamma = G(x + vt)$$

We proceed by two different methods. First, we integrate the PDE in $x \in [0, 1]$ with periodic boundary conditions. There is no loss of generality in considering this domain size as the diffusivity can be changed to vary the spatial scale. If a traveling wave is found then we have a solution of the above form. The second method is to derive a 3-variable ODE by trying the above form and studying the induced flow.

6.1 PDE solution

For $s = 45, e = 0.01$ local kinetics are of the excitable type. Figure 11 shows the traveling pulse solution obtained by integrating in time the initial condition $\gamma(x) = 0.1, \rho(x) = x$. Diffusivity is taken to be 1×10^{-5}. The wave speed is $v \sim -1/83.4$.

Qualitatively, the situation is as follows. Suppose a point at (ρ_0, γ_0) is yanked in the positive γ direction by its neighbor; once excited, locally γ would increase even more before the system returns to equilibrium. This point can in turn excite its other neighbor. The neighbor that excited the point is already in the excited state and does not much get affected. This way, we get a pulse traveling in one direction. The system returns to equilibrium in the wake of the pulse. The periodic boundary condition keeps the pulse in the domain.

We define x_p, the pulse length as the length of the domain where the system is not at equilibrium. If $x_p < 1$ as it is in figure 11 we can construct a family of plane wave solutions by simply separating the pulses by varying amounts. It is possible that $x_p = 1$ as shown in figure 12 where the diffusivity is changed to 1×10^{-3}. Here we do not have a distinct pulse and so it is not possible to generate a family of solutions. Further, there is a period for the oscillations and so we call these periodic solutions.
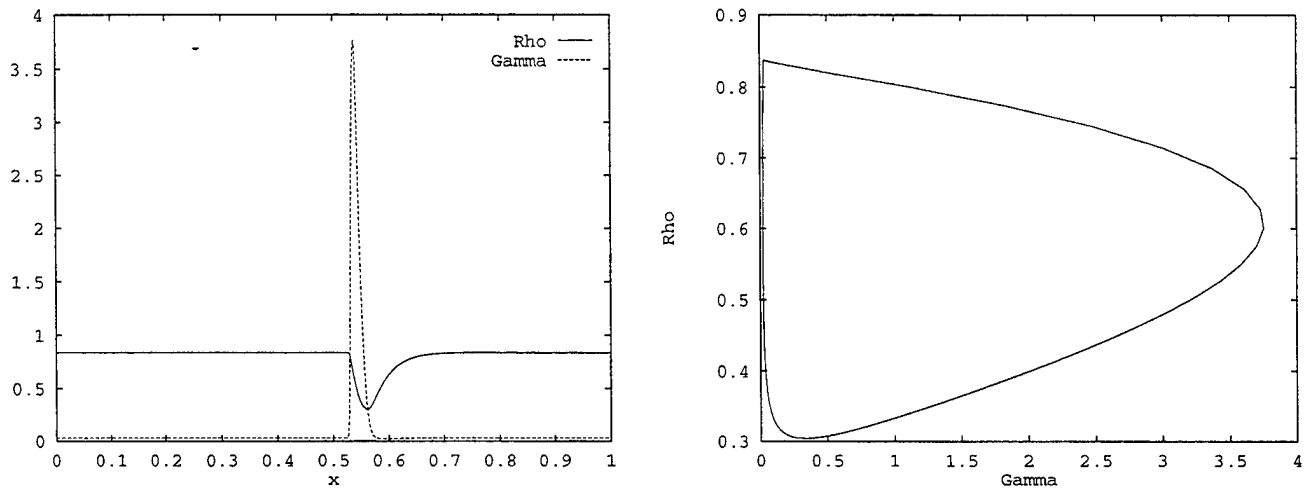
Figure 11: Traveling pulse solution obtained by integrating in time the initial condition $\gamma(x) = 0.1, \rho(x) = x$ in $x \in [0,1]$ with periodic boundary conditions. $D = 1 \times 10^{-5}$. The pulse is moving to the left. The picture on the left shows $\gamma(x)$, $\rho(x)$. On the right are all the points in the domain drawn parametrically in (ρ, γ)-space with lines connecting neighboring points.
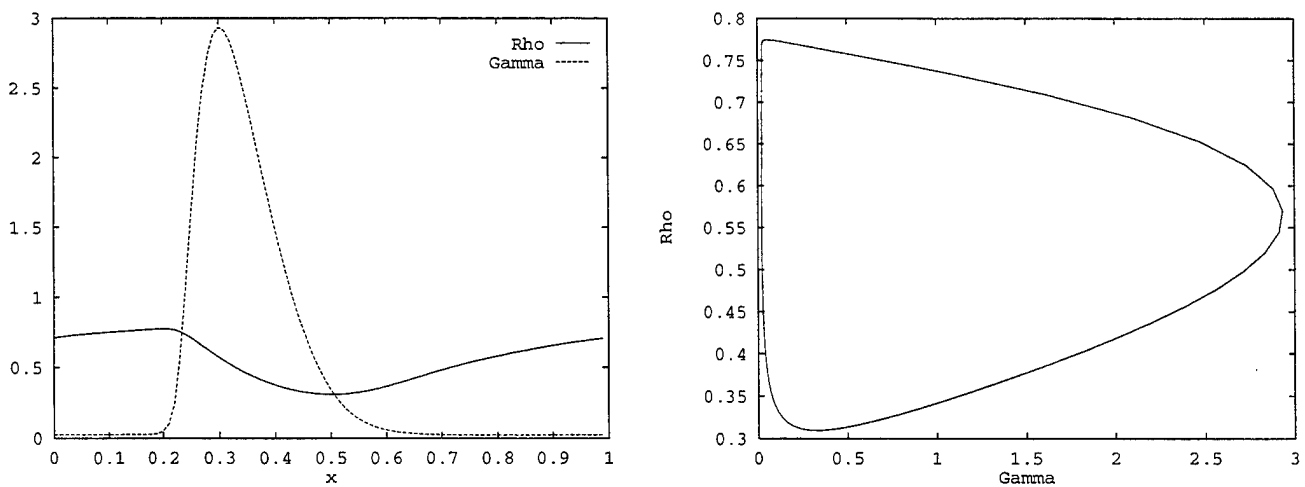
Figure 12: Periodic solution obtained by integrating in time the initial condition $\gamma(x) = 0.1, \rho(x) = x$ in $x \in [0,1]$ with periodic boundary conditions. $D = 1 \times 10^{-3}$. The oscillation is moving to the left. The picture on the left shows $\gamma(x)$, $\rho(x)$. On the right are all the points in the domain drawn parametrically in (ρ, γ)-space with lines connecting neighboring points.

6.2 ODE solution

Inserting $\rho = R(x + vt), \gamma = G(x + vt)$ in (8) and (9), we obtain the following 3-variable ODE :

$$
\begin{aligned}
\frac{dR}{dt} &= \frac{1}{v}\left[-f_1(G)R + f_2(G)(1 - R)\right] \\
\frac{dG}{dt} &= P \\
\frac{dP}{dt} &= \frac{1}{D}\left[vP - \frac{k_1}{e}\left(s\Phi(R, G) - G\right)\right]
\end{aligned}
\tag{14}
$$

The ODE has the fixed point

$$
R_0 = \rho_0, \quad G_0 = \gamma_0, \quad P = 0.
$$

Homoclinic orbits in the induced flow correspond to traveling pulse solutions whereas closed orbits correspond to periodic solutions. Studying this ODE numerically is hard because the spectral radius of the linearization around the fixed point is typically quite large making accurate integrations of the ODE difficult.

In conclusion, we say that solving the PDE directly is a superior method for numerically obtaining plane wave solutions. Having said that, one must also mention that since the PDE picks out only stable solutions, it does not provide complete information. In order to study the pulse speed as a function of domain size and the transition from closed orbits to homoclinic orbits one must investigate the ODE.

7 Circular/target patterns — Role of pacemakers

7.1 1-D analogue

There is a one dimensional analogue of circular/target patterns — a train of pulses emanating from the center. In the previous section, we saw that the model admits a pulse moving in one direction. It would be quite easy to construct a solution where two pulses move away from each other. If no-flux boundary conditions were employed, there would be no mechanism to keep these pulses in the domain and once they have left, the system would return to equilibrium. Experiments with BZ-reaction have suggested that local chemical impurities can serve as pacemakers. It seems likely that this is the case here as well. What could be the mechanism ? One possibility is to imagine the value of s being high locally so that in a small region the local kinetics have a limit cycle behavior whereas the rest of the domain can have excitable behavior. The limit cycle region will act as the pacemaker and be the source of waves which will be propagated through the rest of the domain. Figure 13 shows the result

302

of a simulation with no-flux boundary conditions where $s(x) = 45 + 20\exp\left[-(\frac{0.5-x}{0.05})^2\right]$ in $x \in [0,1]$. $e = 0.01$ and $D = 1 \times 10^{-5}$.
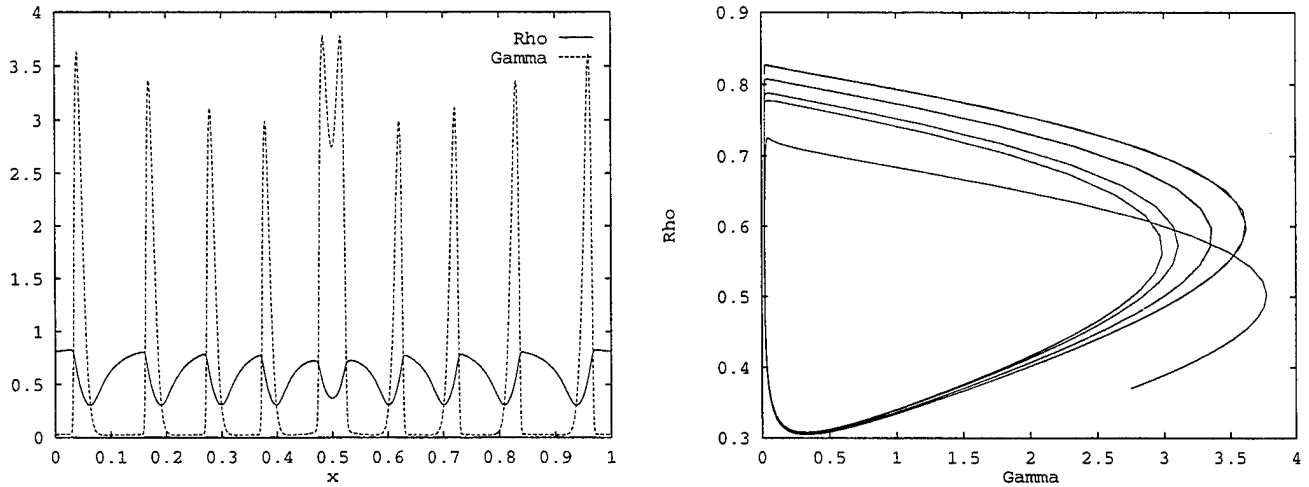
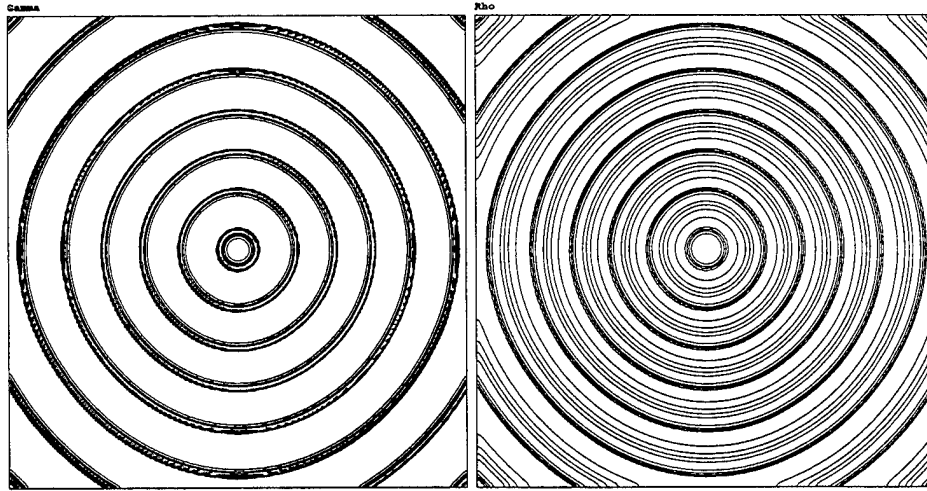Figure 13: 1-d analogue of target pattern obtained by employing an inhomogeneous distribution of $s(x)$. Waves are generated in the center of the domain where the local kinetics have limit cycle behavior. Excitable regions carry the waves to the edge of the domain. $\gamma(x)$, $\rho(x)$ are displayed on the left. On the right are all the points in the domain drawn parametrically in (ρ, γ)-space with lines connecting neighboring points.

7.2 2-D simulations

Figure 14 shows the result of the two-dimensional simulation with the same parameters. It is a radially symmetric version of figure 13.

7.3 Need for pacemakers

It is very easy to get target patterns in 3-variable systems such as the Oregonator. It hence seems likely that the full model with β-dynamics will admit target patterns as well. This is surely related to the fact that inhomogenous distribution of $s(x)$ can create target patterns as well — the distribution of $s(x)$ corrects some of the physics which was removed by assuming that β-dynamics were faster than ρ and γ-dynamics.

Figure 14: Target pattern obtained by employing an inhomogenous distribution of $s(x)$. Waves are generated in the center of the domain where the local kinetics have limit cycle behavior. Excitable regions carry the waves to the edge of the domain. γ is displayed on the left and ρ on the right.

8 Spiral-waves — Genesis and evolution

Spiral waves differ from circular waves in at least two ways. Firstly, there does not seem to be a straightforward one dimensional analogue for spiral waves. Secondly, there is no need for a pacemaker mechanism for spiral waves to form and sustain. Figure 15 shows the formation of a spiral wave. The initial condition is half of a plane wave. The free end curls up to form a spiral wave. Parameters are : $s = 47$, $e = 0.01$, (excitable local kinetics) and $D = 4 \times 10^{-5}$. No-flux boundary conditions are used.

Figure 16 shows the formation of a pair of counter-rotating spiral waves. The initial condition is a segment of plane wave with two free ends. Each free end curls up to form a spiral wave. Notice that when two spiral waves meet, they cancel each other.

Figure 17 shows the interaction of a spiral wave with two pacemakers emitting circular waves. Notice once again that when two waves collide they cancel each other.
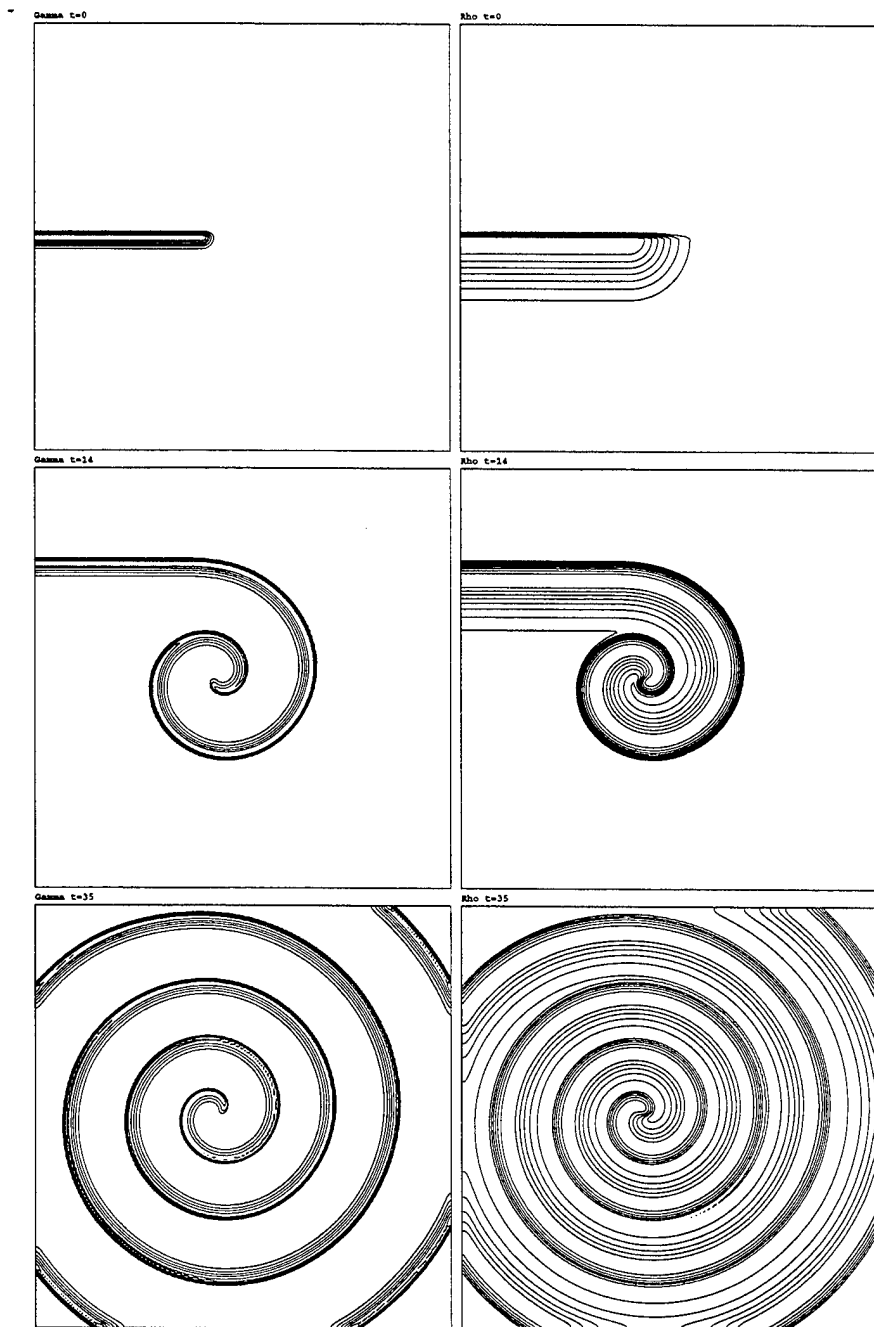
Figure 15: Formation of a spiral wave. γ is displayed on the left and ρ on the right at times $t = 0$, $t = 14$ and $t = 35$.
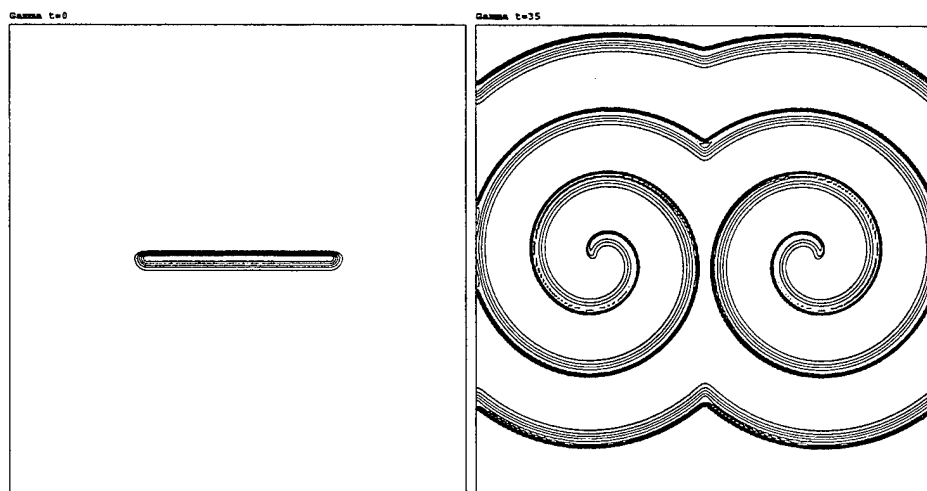
Figure 16: Formation of counter-rotating spiral waves. γ is shown at times $t = 0$ and $t = 35$.
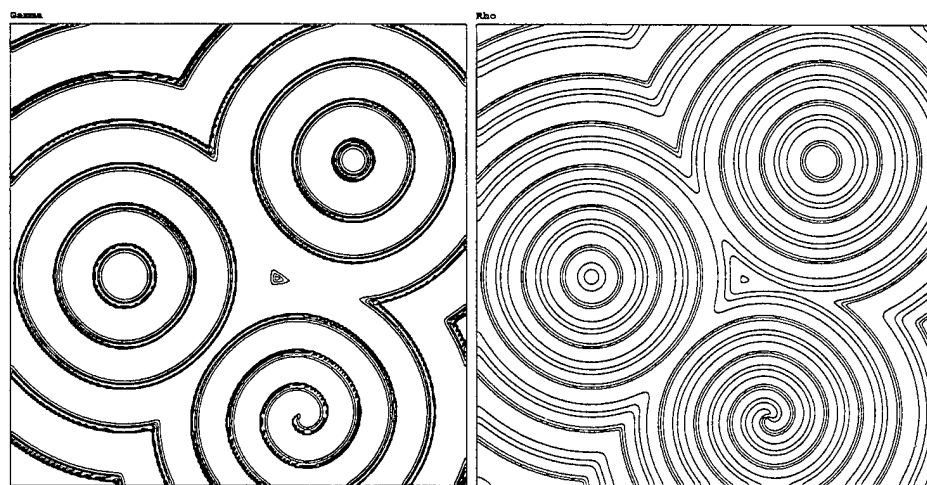
Figure 17: Interaction of spiral and circular waves. γ is displayed on the left and ρ on the right.

9 Evidence for subcriticality

We saw in §5.4 that the complex Ginzburg-Landau equation has local subcritical behavior. Here, we present evidence for subcritical behavior in the full system. Figure 18 summarizes the behavior — starting from an initial condition consisting of perpendicular ramps for ρ, γ, we ask what the final pattern is. For $\delta > 0$, the spatially uniform solution has limit cycle oscillations. This solution is not picked out starting from our initial pattern — the final pattern for $\delta > 0$ is a spiral wave. For $\delta < 0$, however, the situation is different. The spatially uniform solution does not vary in time either and this is the final pattern selected starting from our initial condition. If we follow the spiral wave solution by slowly decreasing δ, we find that a spiral wave solution does exist in the regime with $\delta < 0$ and the way the spiral wave solution is lost is by unwinding of the spiral (figure 19).

Pattern

spiral wave solution

perpendicular ramp IC

spatially uniform
fixed point

spatially uniform
limit cycle oscillations

delta
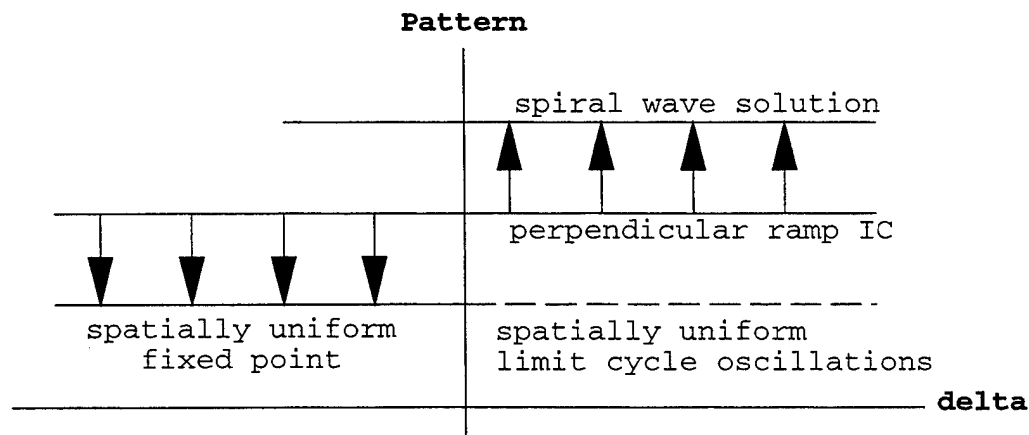
Figure 18: Pattern selection as a function of δ starting from initial condition consisting of perpendicular ramps in ρ, γ.
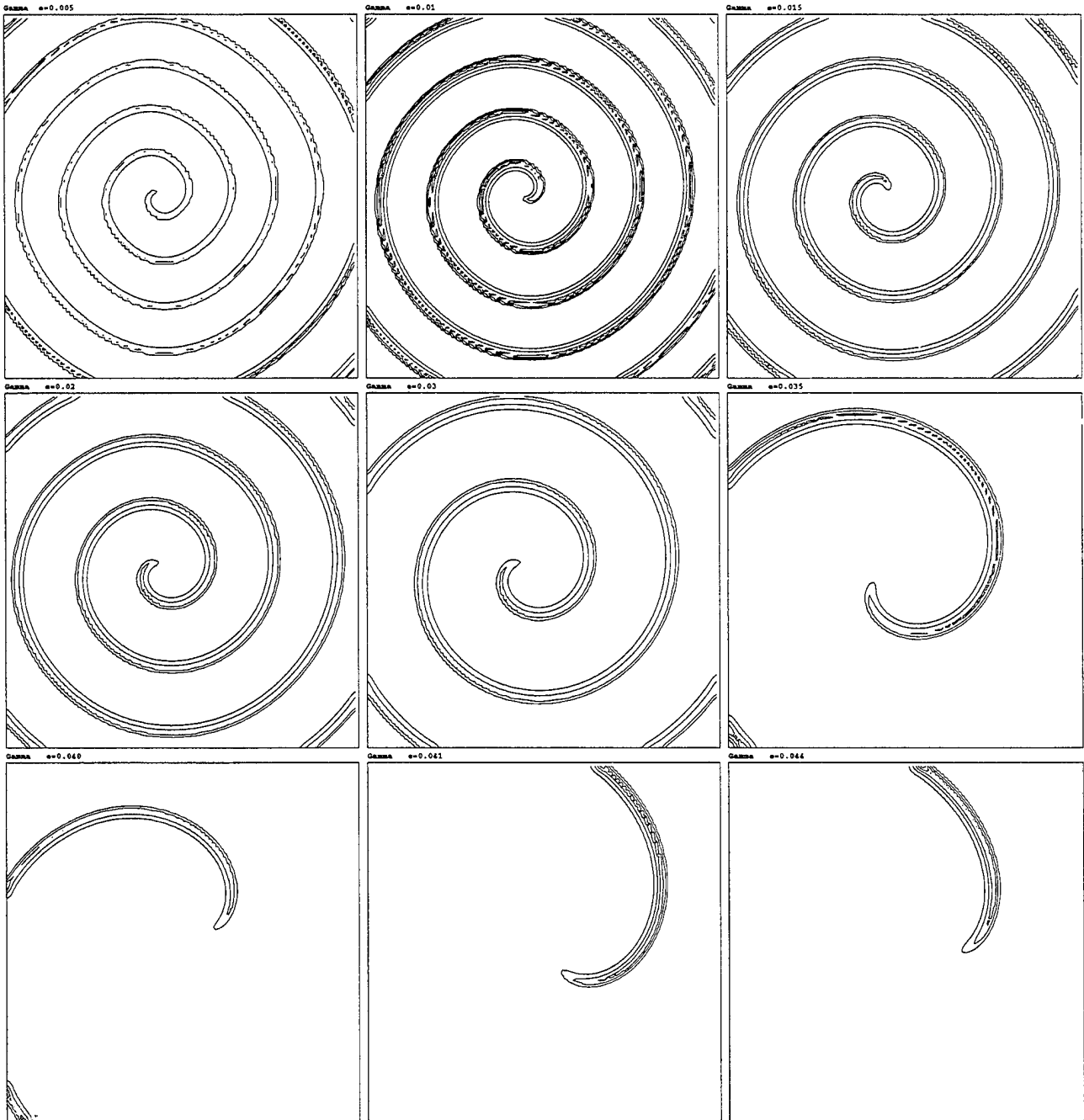
Figure 19: Unwinding of spiral wave as *e* is increased from 0.005 to 0.044.

Acknowledgements

References

[1] CROSS, M. C. & HOHENBERG, P. C. 1993 Pattern Formation outside of equilibrium. *Reviews of Modern Physics* **65**, No. 3, Part II, 851-1112.

[2] KELLER, E. F. & SEGEL, L. A. 1970 Initiation of slime mold aggregation viewed as an instability. *J. Theor. Biol.* **26**, 399-415.

[3] KELLER, E. F. & SEGEL, L. A. 1971 Model for chemotaxis. *J. Theor. Biol.* **30**, 225-234.

[4] KURAMOTO, Y. 1984 *Chemical Oscillations, Waves, and Turbulence* Springer Series in Synergetics, vol. 19, Springer-Verlag.

[5] MARTIEL, J. L. & GOLDBETER, A. 1987 A model based on receptor desensitization for cyclic-AMP signaling in *Dictyostelium* cells. *Biophys. J.* **52**, 807-828.

[6] MURRAY, J. D. 1993 *Mathematical Biology* Biomathematics, vol. 1, Springer-Verlag.

[7] NEWELL, P. C. 1983 Attraction and adhesion in the slime mold *Dictyostelium*, in *Fungal Differentiation : A Contemporary Synthesis*, edited by J. E. Smith, Micology Series 43, Marcel-Dekker Inc., New York, pp. 46.

[8] TYSON, J. J. & MURRAY, J. D. 1989 Cyclic AMP waves during aggregation of *Dictyostelium* amoebae. *Development* **106**, 421-426.

[9] TYSON, J. J., ALEXANDER, K. A., MANORANJAN, V. S. & MURRAY, J. D. 1989 Spiral waves of cyclic AMP in a model of slime mold aggregation. *Physica D* **34**, 193-207.

[10] WINFREE, A. T. 1974 Rotating chemical reactions. *Scientific American* **6**, 82-95.

Topographic Effects on the Propagation of Isolated Eddies

Deborah A. Le Bel

1 Introduction

It has become clear over the past decade that isolated rings can transport large amounts of energy, heat, and salt. Near biogeochemical fronts, such as the Gulf Stream, they can also be important to biochemical fluxes, of both nutrients and organisms (Olson 1990). A better understanding of isolated mesoscale features is thus important to our understanding of the ocean as a bio-physico-chemical system.

There exists a body of literature investigating eddy propagation on a β plane, both numerically (McWilliams and Flierl, 1979; Mied and Lindemann, 1979; Chassignet and Cushman-Roisin, 1991) and analytically (Flierl, 1977; Nof, 1981, 1983a, 1983b; Cushman-Roisin et al. 1990). Briefly, it is noted that an isolated eddy on the β plane will propagate westward near the long Rossby wave speed of the ambient fluid, trailing a Rossby wave wake. Significant nonlinearity damps dispersion and helps to modify this speed; it allows a finite meridional speed, not possible in linear theory. While the distinction must be made between the center of mass and the vortex center as defined by the potential vorticity or displacement extremum (McWilliams and Flierl, 1979; Chassignet, 1989), the latter are more difficult to examine analytically. It has also been shown that under weak dispersion, the zonal drift of the center of mass is actually a robust proxy for that of the vortex center (McWilliams and Flierl, 1979; Chassignet, 1989).

The inclusion of topography is the next logical step in the theoretical investigation of eddies and becomes important when we note that at least two of the energetic ring-producing areas shed vortices which subsequently interact with topography. The Gulf Stream rings pass over the New England Seamounts and may propagate onto the continental shelf (Cheney and Richardson, 1976), and Agulhas rings encounter first the Cape Rise and later the Schmitt-Ott Seamounts (Olson and Evans, 1986). It is must be asked what effects topography has on isolated eddy dynamics.

Topography introduces an effect similar to that of β; in one case, it is the latitudinal variation of the Coriolis parameter that breaks the symmetry around the center of a vortex, and in the other it is the variation of water column height. Both are variations in leading order potential vorticity, $\frac{f}{H}$. The topographic beta effect can be much more complicated than the planetary beta, however. While the planetary effect has a clearly defined direction and a constant amplitude over the space for which the β approximation is valid, the topographic β can be in any direction and of any amplitude. The resulting motion of the eddy cannot be easily predicted for any form other than a simple constant slope.

This work describes the derivation of an expression for the propagation of the center of mass of a vortex over generalized topography. The reduced gravity approximation is used for two reasons. First, there exists a large body of previous work employing reduced gravity (Nof, 1981, 1983a; Chassignet, 1989; Cushman–Roisin et al., 1990; for example). Second, the only two analytically tractable approximations are quasi-geostrophy and reduced gravity. Quasigeostrophy imposes its own limitations on scales, most notably on the size of the interface displacements relative to the layer in which the vortex is embedded.

One drawback to the present approach is that the physical formulation is somewhat unrealistic; most eddies are surface-intensified, while the present one exists in the bottom layer. Nof (1983a) does mention a few observed cases of deep eddies, but they are not the norm. In addition, some processes are excluded using the reduced gravity approximation; barotropic motion and baroclinic instability

(Chassignet and Cushman-Roisin (1990)), specifically. No form drag is introduced by interaction with a dynamically active upper layer; this can become important because form drag initially opposes the motion of the vortex, it tends to slow the vortex zonally and induce further meridional motion, much as the form drag induced by the Rossby wave wake does. Additionally, with the barotropic-equivalent formulation, vertical modes are limited to one and modal coupling is restricted to self-interacting.

Smith and O'Brien (1983) investigated the behavior of a β-plane vortex in a two layer, primitive equation model with topography meant to simulate the continental slope – a down-slope meeting a flat bottom. They found that a layer thickness ratio of 0.25 was sufficiently small to insulate an upper-ocean eddy from topographic influences, while a larger ratio of 0.66 showed some transfer of energy from the baroclinic to the barotropic mode, as was seen by McWilliams and Flierl (1979) for a flat bottom. Barotropic anticyclones developed an eastward propagating vortex pair similar to the modon in the flat-bottomed experiment, while cyclones continued to propagate onto topography, evolving into an upper-ocean vortex pair; thus there was some evidence that cyclones will be preferentially trapped on-shore and indicated the possibility of greater persistence of anticyclones due to the less dispersive nature of the eastward-propagating vortex pair. In contrast to the flat-bottomed case (McWilliams and Flierl, 1979), some development of baroclinicity was forced. Of particular interest is Smith and O'Brien's results with lower-layer eddies. Behavior was much like the barotropic case; anticyclones evolved to modon-like structures, and cyclones moved eastward onto the topography and evolved to upper-ocean structures. Self-advection speeds were significantly larger than in the barotropic case, however.

The reduced-gravity system results in a barotropic-equivalent vortex. Chassignet and Cushman-Roisin (1991) explored the applicability of the reduced gravity approximation and found empirically that a ratio of upper to lower layer depths $\frac{H_1}{H_2} \leq \frac{1}{50}$ was sufficiently small to legitimatize the use of reduced gravity dynamics. A scaling analysis more formally specified

$$ H \gg \max \left[H_1, \ \min(H_1 \frac{L^4}{R_d^4}, \ h_o \frac{L_\beta}{L}) \right] $$

where H is the total depth and L_β is the planetary scale. It easy to see that the model used in this study did meet these requirements. Unfortunately, it is clear that an unreasonably deep upper layer is required - in this case, at least 50 km.

The layout of the paper is as follows. In section 2, the derivation of an expression for the propagation of the center of mass of a vortex over generalized topography is presented. An approximation of constantly slope is then made and the results validated with numerical simulations in section 3. Some unpredicted behavior of the center of mass will also be discussed. Higher order calculations are presented in section 4 to try to explain that behavior. Finally, conclusions are summarized in section 5.

2 Derivation of the Center of Mass Translation

The derivation of an expression for the translation of the center of mass follows that of Cushman-Roisin et al. (1990). The equations of motion for a one and a half layer, reduced gravity system are

$$ u_t + u u_x + v u_y - fv = -g' \eta_x $$

$$ v_t + u v_x + v v_y + fu = -g' \eta_y $$

$$ h_t + \nabla \cdot (h\mathbf{u}) = 0 $$

where

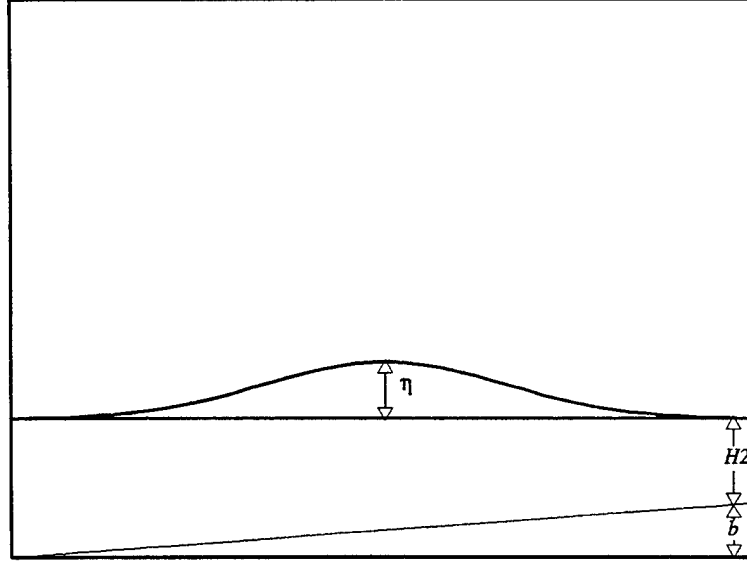$$ h = H_2 + \eta - b $$

311

Figure 1: Diagram of the vertical structure of the theoretical and numerical model.

A schematic of the system is shown in Figure 1. H_2 is the undisturbed bottom layer depth, η the fluctuating interface between the two layers, which defines the vortex, and b the topography. The β-plane approximation is made, so that the Coriolis parameter is given by f $=$ $f_0 + \beta y$.

The equations are scaled by L, a length scale not constrained to be the Rossby radius; δH, an interface displacement scale not constrained to be $\ll H_2$; and T, a time scale not constrained to be the advective or inertial time scale, to give

$$\omega u_t + \epsilon u u_x + \epsilon v u_y - v - \beta y v = -\eta_x$$

$$\omega v_t + \epsilon u v_x + \epsilon v v_y + u + \beta y u = -\eta_y$$

$$\omega \eta_t + s\nabla \cdot \mathbf{u} + \epsilon \nabla \cdot (\mathbf{u}\eta) - \gamma \nabla \cdot (\mathbf{u}b) = 0$$

where

$$\omega = \frac{1}{f_0 T} \qquad \text{time scale}$$

$$s = \frac{g'H_2}{f_0^2 L^2} \qquad \text{Burger (stratification) number}$$

$$\epsilon = \frac{g'\delta H}{f_0^2 L^2} \qquad \text{Rossby number}$$

312

$$\gamma = \frac{g'b}{f_o^2 L^2} \qquad \text{topographic effect}$$

$$\beta = \frac{\beta_o L}{f_o} \qquad \text{beta effect}$$

The order of the error is given by

$$\max(\omega, s, \epsilon, \gamma)$$

and specifies which effect dominates the dynamics at the next order. Strictly speaking, the Burger, Rossby, and topographic numbers given above are truly squared values of the actual ratios.

Noting that the largest term containing ω must be of the same order as the largest term not containing ω, and that the scaling of H was by $\delta H \leq H$, requires $\epsilon \leq s$, ω is given by

$$\frac{\max(s\epsilon, s\beta, s\gamma)}{\max(1, s, \gamma)}$$

and the order of the error is reduced to

$$\delta = \max(\epsilon, \beta, \gamma)$$

for cases with the vortex larger than the radius of deformation (s <1).

Next, the center of mass is defined as

$$X = \langle x\eta \rangle$$

$$Y = \langle y\eta \rangle$$

$$\text{where } \langle \bullet \rangle = \frac{\iint \bullet \ dx\, dy}{\iint \eta \ dx\, dy} \qquad \text{over an infinite domain}$$

Differentiating yields

$$\frac{dX}{dt} = \langle x\eta_t \rangle$$

$$\frac{dY}{dt} = \langle y\eta_t \rangle$$

Substituting for η_t from above and integrating by parts, the result is

$$\omega \frac{d^2 X}{dt^2} - \frac{dY}{dt} = \frac{s\beta}{\omega}\langle yv \rangle + \frac{\epsilon\beta}{\omega}\langle y\eta v \rangle - \frac{\beta\gamma}{\omega}\langle byv \rangle \frac{\gamma}{\omega}\langle b_x \eta \rangle -$$

$$\omega \frac{d^2 Y}{dt^2} + \frac{dX}{dt} = -\frac{s\beta}{\omega}\langle yu \rangle - \frac{\epsilon\beta}{\omega}\langle y\eta u \rangle - \frac{\beta\gamma}{\omega}\langle byu \rangle \frac{\gamma}{\omega}\langle b_y \eta \rangle +$$

Geostrophic velocities when substituted yield

$$\omega \frac{d^2 X}{dt^2} - \frac{dY}{dt} = -\frac{\gamma}{\omega}\langle b_x \eta \rangle - \frac{\beta\gamma}{\omega}\langle b_x y \rangle + O(\frac{s\beta\delta}{\omega}) \tag{1}$$

$$\omega \frac{d^2 Y}{dt^2} + \frac{dX}{dt} =$$

$$-\frac{\gamma}{\omega}\langle b_y \eta \rangle - \frac{\beta\gamma}{\omega}\langle b\eta \rangle - \frac{s\beta}{\omega}\langle \eta \rangle - \frac{\epsilon\beta}{2\omega}\langle \eta^2 \rangle - \langle b_y y\eta \rangle + O(\frac{s\beta\delta}{\omega}, \omega^2) \tag{2}$$

The formal error is on the order of $[\max(\frac{\beta_o g'}{f_o^2}, \frac{g'b}{f_o})] \times [\max(\delta, \frac{\omega^3}{s\beta})]$.

313

The results obtained without topography (Cushman-Roisin et al., 1990)

$$\omega \frac{d^2 X}{dt^2} - \frac{dY}{dt} = -O(\frac{s\beta\delta}{\omega})$$

$$\omega \frac{d^2 Y}{dt^2} + \frac{dX}{dt} = -\frac{s\beta}{\omega}\langle\eta\rangle - \frac{\epsilon\beta}{\omega}\langle\eta^2\rangle + O(\frac{s\beta\delta}{\omega})$$

are recovered for b = 0 in (1) and (2).

Reconverting to dimensional quantities and assuming that ω is small enough to justify dropping the second derivatives (which it must be to justify a geostrophic scaling), the final expression is:

$$\frac{dX}{dt} = \frac{-\frac{g'}{f_o} \iint \{\eta[b_y + \frac{\beta}{f_o}(H_2 + yb_y + b)] + \frac{\beta\eta^2}{2f_o}\} \; dx \; dy}{\iint \eta \; dx \; dy} \tag{3}$$

$$\frac{dY}{dt} = \frac{\frac{g'}{f_o} \iint b_x\eta(1 - \frac{\beta y}{f_o}) \; dx \; dy}{\iint \eta \; dx \; dy} \tag{4}$$

A few things are important to note here. First, in the f-plane case, the solution is linear and recovers that of Nof (1983a). Nof's solution was for a lens sitting directly on topography of constant, linear slope and included an assumption of constant drift speed, which was not made here. Thus, Nof's solution is applicable under a wider range of conditions than those for which it was derived.

The zonal drift speed is proportional to the meridional slope of the topography and is independent of the intensity of the eddy:

$$\frac{dX}{dt} = -\frac{g'b_y}{f_o}$$

$$\frac{dY}{dt} = 0$$

It is interesting to note that the planetary β introduces a nonlinearity which gives the solution a theoretical dependence on the eddy intensity or amplitude, while the topographic β, even when linked to the planetary β, has no such nonlinearity nor dependence.

Second, a non-zero zonal derivative of the topography will induce a meridional drift; however, in the absence of a zonal derivative, the meridional drift to leading order is zero. Contrast this to the results of Cushman-Roisin et al. (1990), where to first order the meridional drift on a β-plane was always zero. Thus, topography may introduce a finite drift in the y-direction.

Before going on to the validation of the analytical expression by numerical simulation, it is instructive to examine more closely the terms in the zonal drift equation. Assuming topography with a constant meridional slope, it becomes:

$$\frac{dX}{dt} = -\frac{g'\beta H_2}{f_o^2} - \frac{\beta}{2f_o}\langle\eta^2\rangle - \frac{2g'\beta b_y}{f_o^2}\langle y\eta\rangle - \frac{g'}{f_o}b_y$$

$$\text{(a)} \qquad \text{(b)} \qquad \text{(c)} \qquad \text{(d)}$$

Term a is the long Rossby wave speed for the ambient fluid and is always westward. This is also the quasi-geostrophic limit (McWilliams and Flierl, 1979).

Term b will be negative for anticyclones ($\eta > 0$) and positive for cyclones.

Term c couples β and the topography and depends on the sign of the meridional derivative of the topography.

Term d depends on the direction of the slope and thus may enhance or lessen the Rossby wave speed. Again, it is independent of eddy intensity and captures the results of Nof (1983a).

The size of the drift speed predicted by the center of mass derivation is estimated by substituting a Gaussian form for η with a 300 m initial amplitude is substituted into the above analytical expression. The dominant effect, as also noted in the flat bottom case of Cushman-Roisin et al. (1990), is the long Rossby wave speed at -4 cm s^{-1}. Also important is the term involving only the slope, which adds another -2 cm s^{-1}. The η^2 term contributes another .4 cm s^{-1}, again with the sign depending on the sign of the interface displacement. The last term, coupling β and topography, is insignificant, on the order of 10^{-7} cm s^{-1}. Thus, theory predicts that cyclones with the above height perturbation move at approximately -5.6 cm s^{-1} and their corresponding anticyclones at -6.4 cm s^{-1}.

3 Model Description

In order to test this theoretical result, a series of numerical simulations of an isolated vortex over sloping topography were performed. The Burger number, comparing the Rossby radius with the size of the vortex, was held constant while eddy intensity and bottom slope were varied. f-plane and a few simple β-plane experiments were conducted. Previous work has also examined the velocity contrast between layers as a measure of the baroclinicity of the vortex (Mied and Lindemann, 1979; Smith and O'Brien 1981); with the reduced gravity approximation, the motion remains barotropic- equivalent. The model used in this study was an adiabatic, unforced, one and a half layer version of the Miami Isopycnic Coordinate Ocean Model (MICOM), also known as the Bleck-Boudra model (1986). This model has already been used in a series of vortex experiments (Chassignet, 1989; Chassignet and Cushman- Roisin, 1991; Chassignet et al., 1990)

Briefly, the model domain was 2000 x 2000 km, with a uniform grid spacing of 20 km. The lower layer depth, H$_2$, was 1000m, with the ratio of upper to lower layer depth taken to be one thousand in order to practically approximate infinity, in accordance with reduced gravity dynamics. Free slip boundary conditions were used, although the formulation becomes unimportant for this problem; it is the behavior of the isolated feature free of other influences which is under investigation, and the dynamics become significantly different once the boundaries begin to influence the vortex. The initial topography was a simple meridional slope of 10^{-4}, shallowing to the north. A Laplacian friction parameterization was used with a coefficient of $200 \times$ m^2 s^{-1}.

The time step was twenty minutes, with a typical integration time of 200 days. Each integration was initialized with a Gaussian height profile $h = h_o e^{-\frac{r^2}{2L^2}}$ centered in the domain at (1000 km, 1000 km). Here h$_o$, the maximum displacement, varies from 200 to 500; L, the radius of maximum velocity, is 50 km; and r is the distance from the center. The Rossby radius of deformation is 45 km.

The scaling for the numerical eddy is compared to typical values for open ocean rings, Gulf Stream rings, and intrathermoclinic eddies in Table 1. Based on Rossby number, none of these features can be classified as strictly geostrophic, although the numerical case clearly deviates the farthest. The gentle slope incorporated in the model gives a topographic β effect on the same order as the nondimensional planetary β effect experienced by the numerical ring.

An alternate comparison may be made for the position in Rossby number— Burger number space. If we assume the flow is in gradient balance

$$\frac{v^2}{r} + fv = g'\frac{\partial h}{\partial r}$$

we can scale the equation as

$$\epsilon^2 + \epsilon = s$$

using

$$\epsilon = \frac{v}{Lf} \qquad s = \frac{g'\delta h}{f^2 L^2} \qquad v = \text{maximum azimuthal velocity}$$

$$L = \text{radius of maximal velocity}$$
$$\delta h = \text{interface change across ring}$$

This scaling for the Burger number corresponds to the Rossby number in the center of mass theory derivation. In either case, the Rossby number measures the magnitude of the flow speed relative

Table 1: Scaling

			Eddy	Open Ocean	Gulf Stream	Intrathermoclinic
g'			.02	.02	.02	.01
L			50 km	50 km	100 km	30 km
H	-		1000 m	400 m	500 m	50 m
δH			300 m	400 m	500 m	50 m
ϵ	Rossby number	$\frac{g'\delta H}{f_0^2 L^2}$	.24	.08	.10	.11
s	Burger number	$\frac{g'H}{f_0^2 L^2}$	.80	.30	.10	.11
γ	Topography	$\frac{g'b}{f_0^2 L_b^2}$	.02			
β	Beta effect	$\frac{\beta_0 L}{f_0}$	.01	.01	.02	.006
ω	Time scale	$\frac{1}{\omega T}$	.192			

a typical advective speed set by the vortex length scale and the inertial period; in the theoretical derivation, the Burger number compares the "external" deformation radius with the vortex length scale, while in the present case it compares the "internal" deformation radius.

This scaling gives

$$\epsilon = -\frac{1}{2} \pm -\frac{1}{2}(1 + 4s)^{\frac{1}{2}}$$

with + corresponding to cyclones and − to anticyclones. Note there is a solvability bound at s=−.25; there is a limit to where the pressure gradient can be balanced by a positive centripetal force. This bound means that cyclones may be more nonlinear and more energetic than their anticyclonic counterparts. Figure (2) shows the scaling of many observed rings in addition to that of the 200 m, 300 m, and 500 m Gaussian eddies used in this study. For this particular choice of reduced gravity, a 500 m anticyclone cannot exist in gradient balance, while the 300 m is marginally stable. Some adjustment to a more stable configuration, as observed in oceanic rings, may occur within the numerical model.

3.1 Comparison of Analytical and Numerical Results

It was elected to initialize velocities in gradient balance rather than pure geostrophic balance, as it was anticipated that the adjustment process would be gentler. There are two processes at work; the first is the growth of the centripetal acceleration term to develop gradient balance, and the second is the adjustment to the bottom slope from a flat-bottomed (symmetric) initialization. An adjustment similar to this second occurs on a flat-bottomed β-plane, which is in general initialized with f-plane (again symmetric) geostrophy. While it would be preferable to start the vortex in full gradient equilibrium, thus also eliminating any adjustment effects from the topography, results of an experimental comparison did not significantly differ between the two versions of gradient balance. Vortex amplitude changes differed only over the first ten days, and zonal drift speeds were essentially identical.

Figure 3 shows the evolution of the interface perturbation of a 300 m anticyclone on an f-plane over the standard 1×10^{-4} slope. The time interval between panels is 40 days and spans t=0 to t=200 days. The behavior of this particular vortex is quite typical; the anticyclone drifts southwestward, decaying as it goes. By the end of the model run, the displacement extremum has been reduced to 125 m, and the Rossby wave wake, which has appeared in the pressure field by day 40, has obtained a maximum amplitude of 15 m. Also, an extension of the leading edge of the vortex is seen, so that it is no longer axisymmetric.

The first set of experiments involved evaluating the analytical prediction of intensity independence by comparing a set of model runs with initial amplitudes of 200 m, 300 m, and 500 m. Velocities
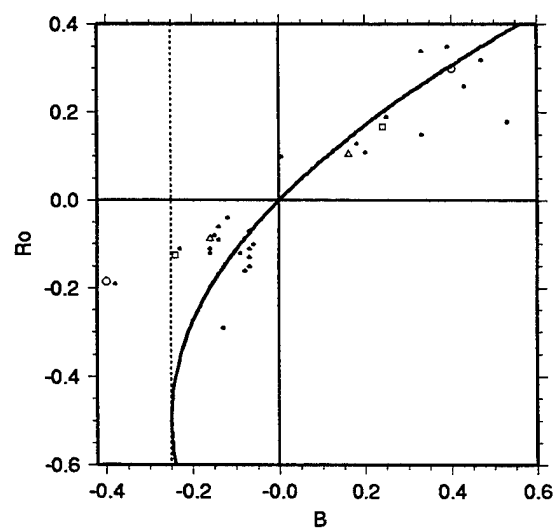
Figure 2: Distribution of Rossby (R_o) and Burger numbers (B) of thirty-five observed rings from several different energetic regions. Also included are values for vortices examined in this study. $\triangle$ 200 m, $\square$ 300 m, $\circ$ 500 m. Re-drawn from Olson (1991).
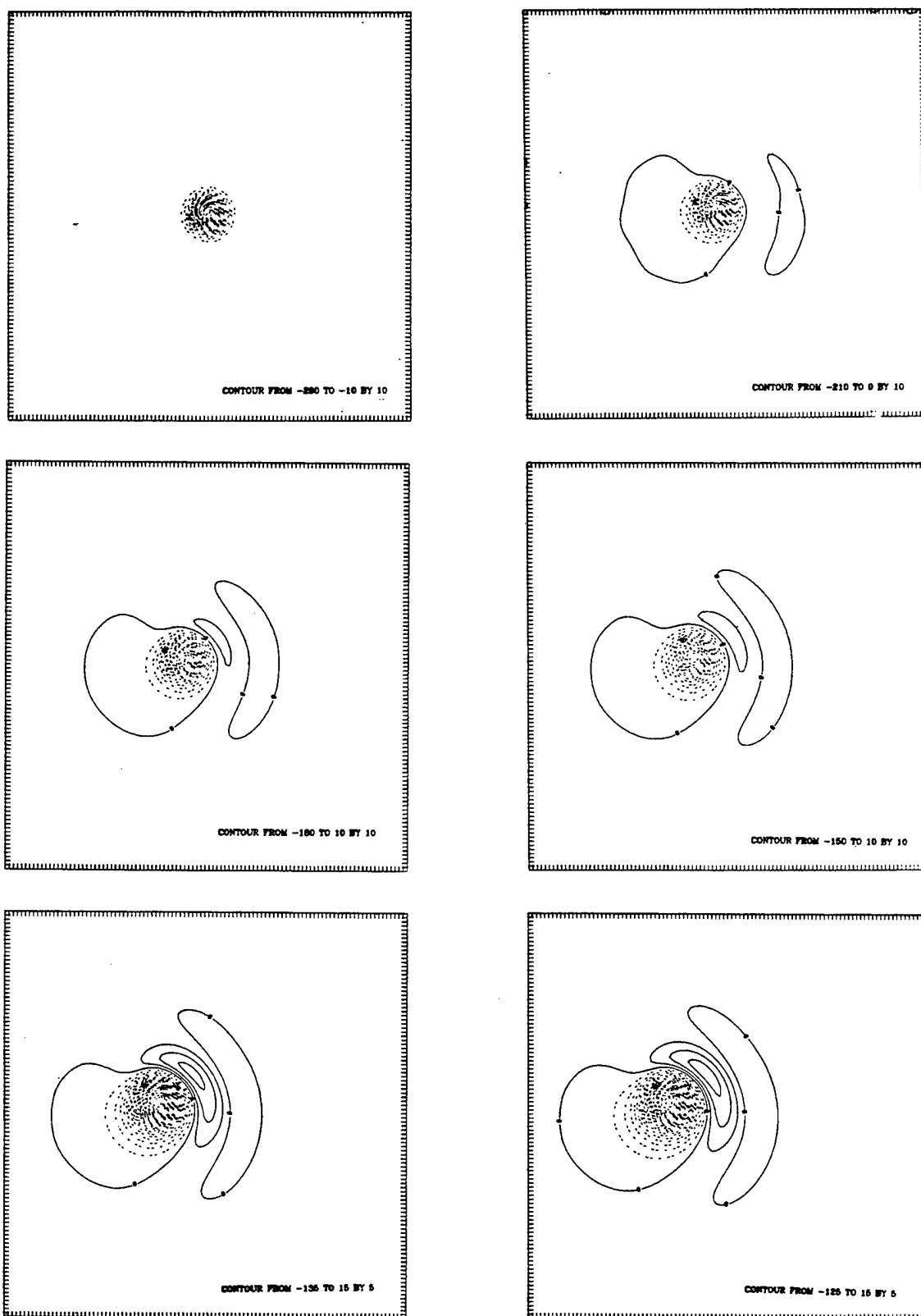
Figure 3: Evolution of interface pressure contours for a 300 m anticyclone on an f-plane with the standard bottom slope. Time interval is forty days.

were calculated in geostrophic balance for this set of experiments, as, again, a 500 meter anticyclone in gradient balance is not a viable solution with this choice of parameters.

Figure 4 presents the zonal drift speeds for the three cases, both cyclonic and anticyclonic in circulation. In addition, both center of mass and height extremum estimates are shown. Initial center of mass velocities are indeed independent of amplitude but begin to diverge by day 30 for the cyclones and day 60 for the anticyclones (possibly reflecting an adjustment process, which is expected to be stronger for cyclones). Cylcones slow, and anticyclones speed up, with the more linear (shallower)vortices deviating farther from the predicted -2 cm s^{-1}. Drift speeds for the height maxima are approximately -1.8 cm s^{-1}, close to the cyclonic center of mass value, and show no differentiation based on the sign of circulation. The degree of independence from amplitude is also greater for the height maxima. Finally, the most salient feature of this figure is the oscillation with an apparent period of twenty days which occurs in both center of mass estimates. It is not as apparent in the height maximum estimates, although a hint does appear within the 500 m anticyclone.

Meridional velocities (fig. 5) show the same qualitative behavior. Again, the 20 day oscillation appears in the center of mass speeds; the motion is almost purely oscillatory, although a small average speed of about ± 0.03 cm s^{-1} is maintained. The more linear vortices again show larger amplitude. The height maximum estimates, an order of magnitude higher than those of the center of mass, decay to a value of ± 0.2 cm s^{-1}, reminiscent of the finding of McWilliams and Flierl (1979) that a strongly nonlinear vortex will eventually decay to the point where linear dynamics once more apply, with a zero meridional drift and a Rossby wave speed. Again, any oscillatory motion is not clearly expressed.

Trajectories are shown in figure 6. What is clear here is that both the center of mass and displacement extremum show less meridional motion and stronger zonal propagation with increasing displacement (nonlinearity). The height maximum estimates show no difference in the propagation speed of cyclones and anticyclones, while the center of mass of cyclones clearly move less quickly than do that of anticyclones. The 200 and 300 m cyclones show less oscillation than do their corresponding anticyclones.

Figure 7 shows the translation speeds for a 300 m cyclone and anticyclone over a doubled slope of 2×10^{-4}. While the center of mass zonal drift initially matches the predicted value of -4 cm s^{-1}, both vortices show reduced speed, with the cyclone translating at about -2.2 cm s^{-1} and the anticyclone at -3.2. The displacement extrema, however, both start at near -2 cm s^{-1} and asymptote to -3.5 cm s^{-1}, again with no differentiation between cyclone and anticyclone. Both meridional drift estimates are approximately doubled. At the end of the experiment, both centers of mass are moving poleward; the self-advection has increased in speed so that the vortices have actually impinged upon the western wall and begun to slide up-slope. The last sixty days of these experiments do not exhibit true isolated vortex dynamics. The appropriate average speeds are $-4 \pm .2$ cm s^{-1}.

These two sets of experiments confirm the analytical prediction for the center of mass translation on an f-plane; results depend linearly on the topographic slope, and results are independent of eddy amplitude. While differences do exist, it is hypothesized that these are due to interaction with the Rossby wave wake. It should also be pointed out that the error on the analytical expression is quite large, about 24%. This is driven mainly by the ω term, which if not small produces an inertial oscillation. This oscillation is superimposed on the translation but should not affect it. Thus the form error of the drift speed prediction is an upper bound for the "true" one.

A list of the experiments performed, along with average and initial drift speeds, is included in Table 2.

At present, there are four possible explanations for the oscillatory behavior of the center of mass:

(1) Inertial Oscillation

Because the ω term is not small, the exclusion of the second derivative is not justified. However, there is no apparent physical mechanism for exciting inertial oscillations. Cushman-Roisin et al. (1990) point out that an initial condition with net momentum is necessary to excite an inertial oscillation. In addition, twenty days is not close to the inertial period.

(2) Interaction with the Rossby wave wake, including a beating with the inertial period.
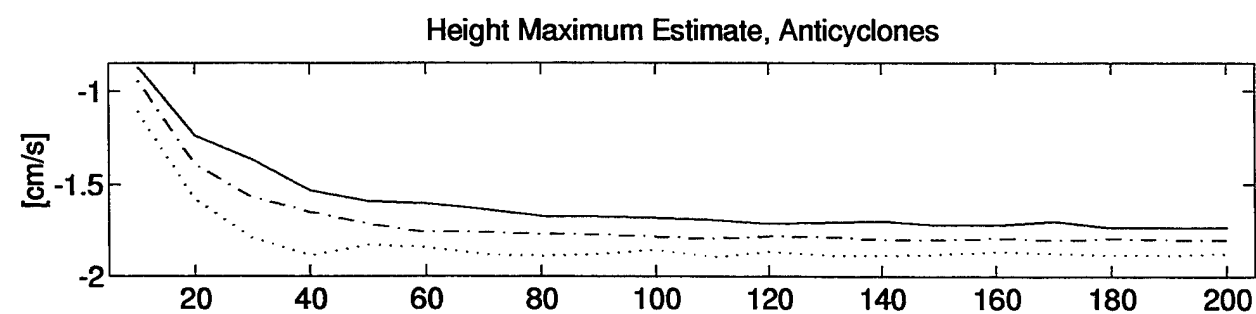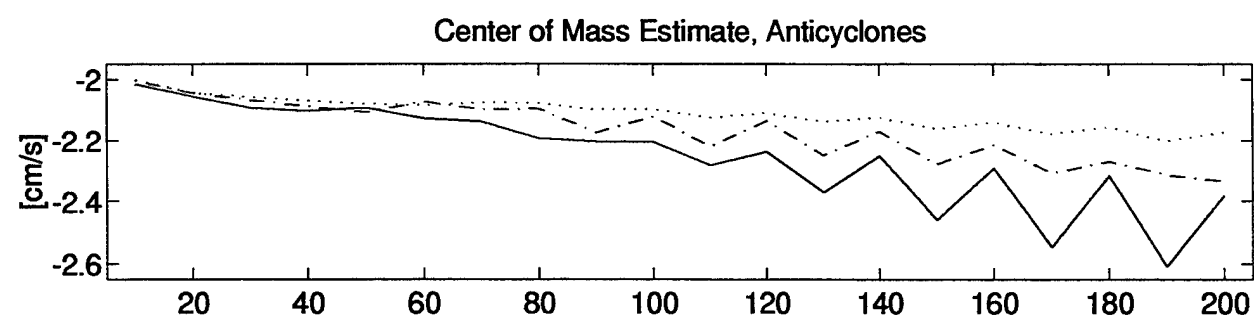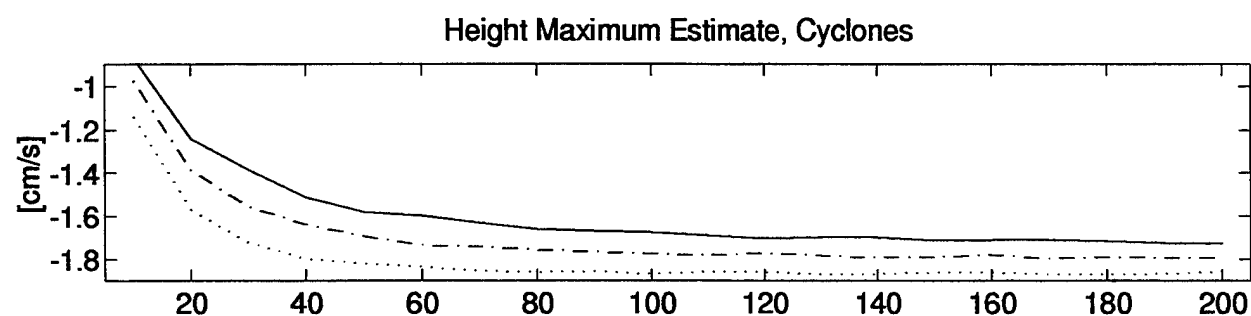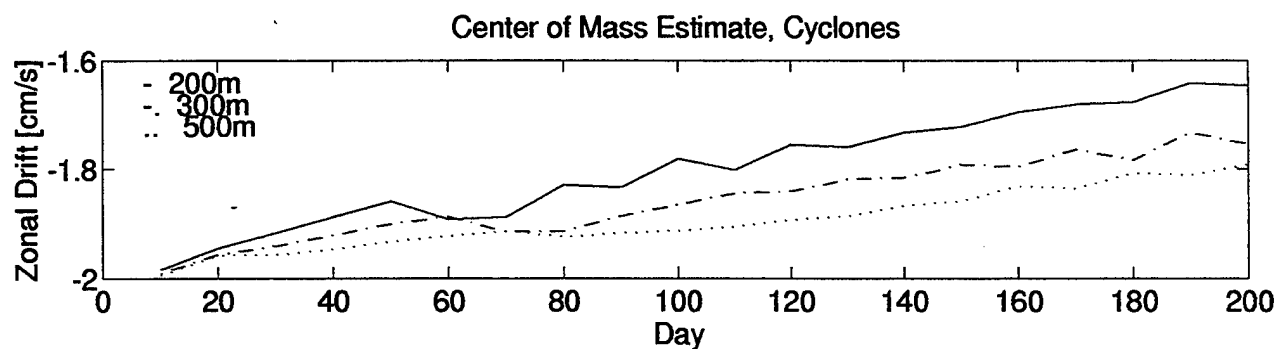
(3) An instability process

319

Figure 4: Zonal self-advection speeds for the center of mass and height extrema for 200 m (solid line), 300 m (dot-dashed line), and 500 m (dotted line) cyclones and anticyclones.
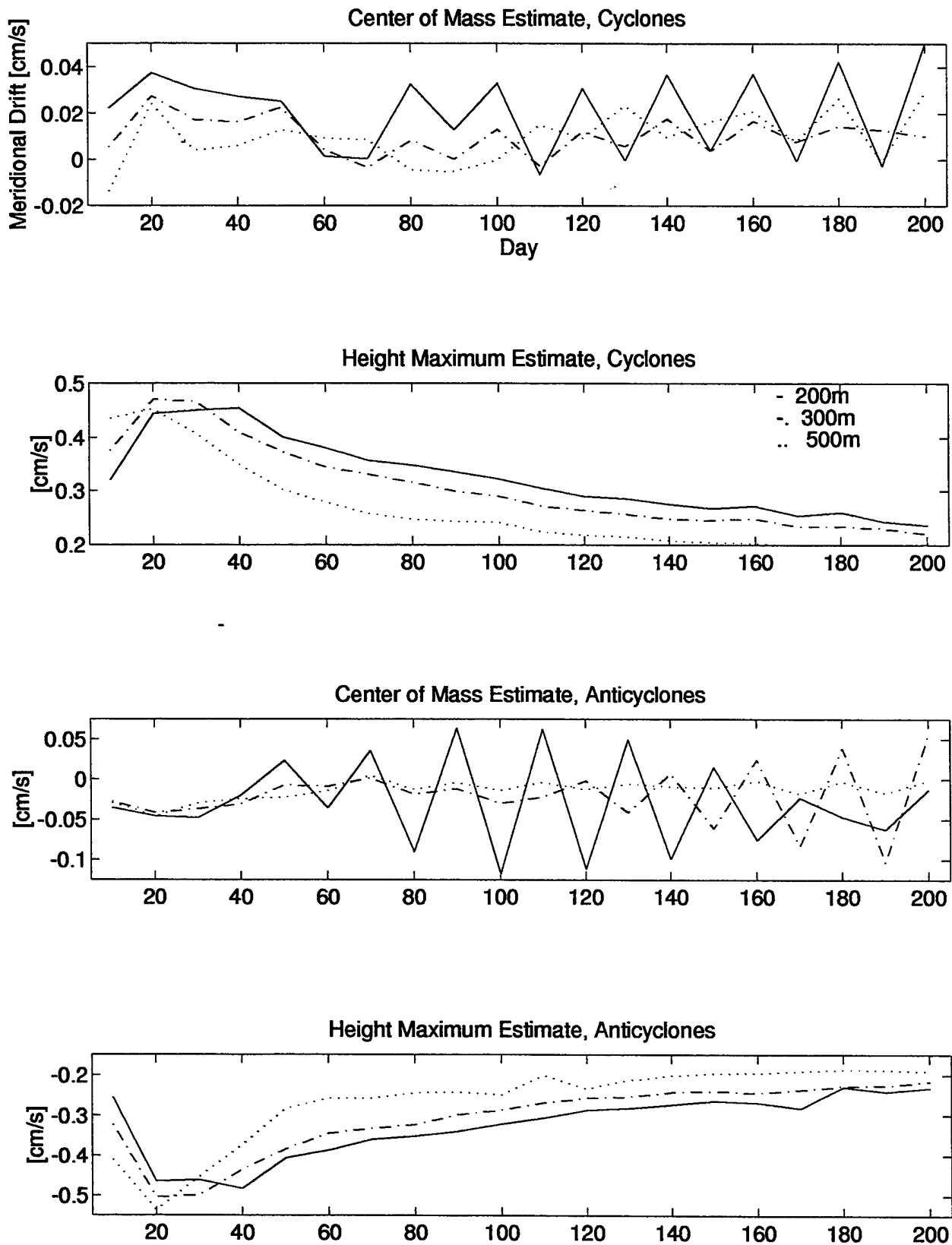
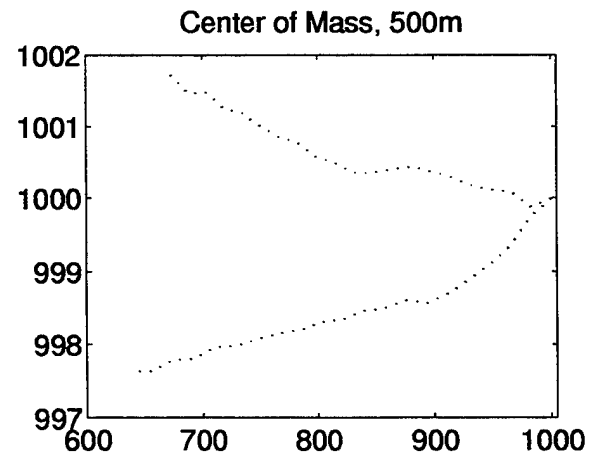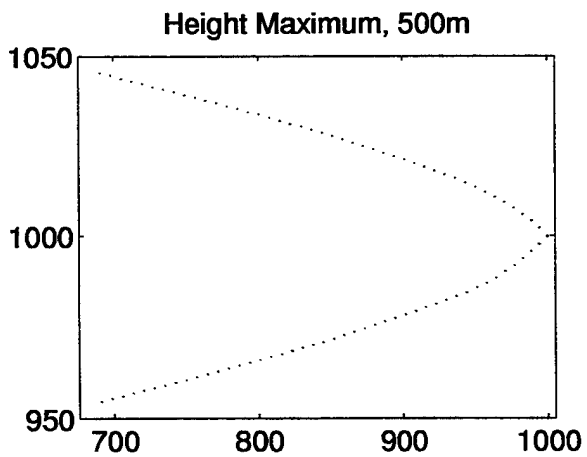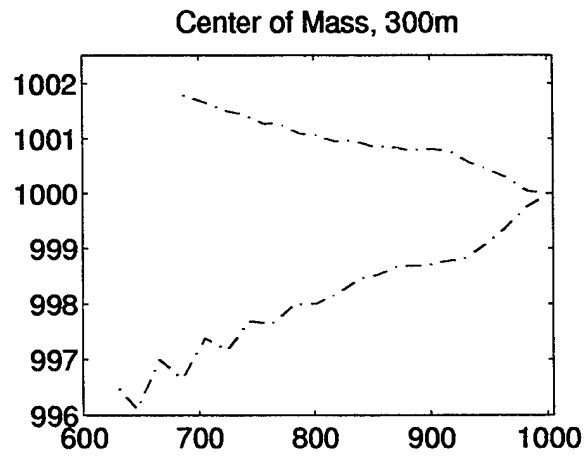Figure 5: Meridional self-advection speeds, as in Figure 4.

Height Maximum, 200m

Center of Mass, 200m

Height Maximum, 300m

Center of Mass, 300m

Height Maximum, 500m

Center of Mass, 500m
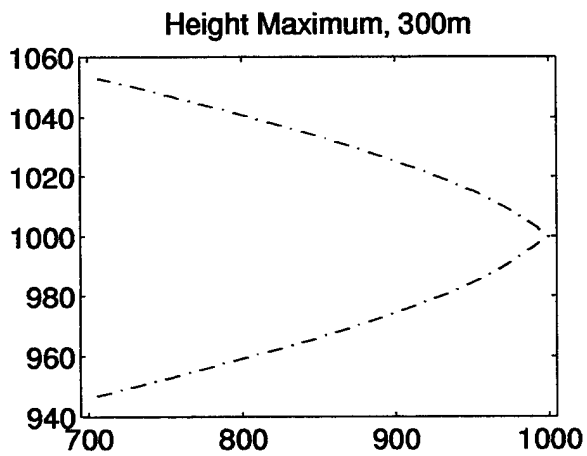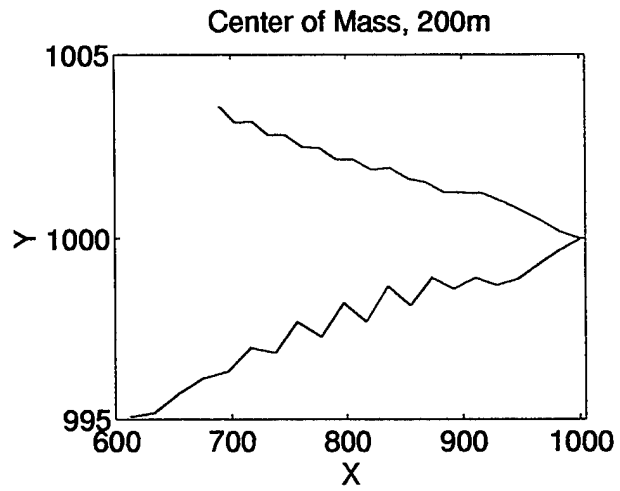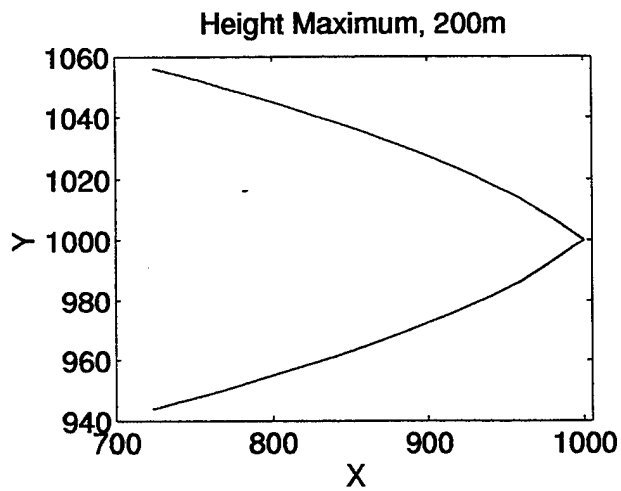
Figure 6: Trajectories of 200 m (solid line), 300 m (dot-dashed line), and 500 m (dotted line) cyclones and anticyclones on an f-plane and standard sloping topography. Cyclones move poleward and anticyclones equatorward.

322
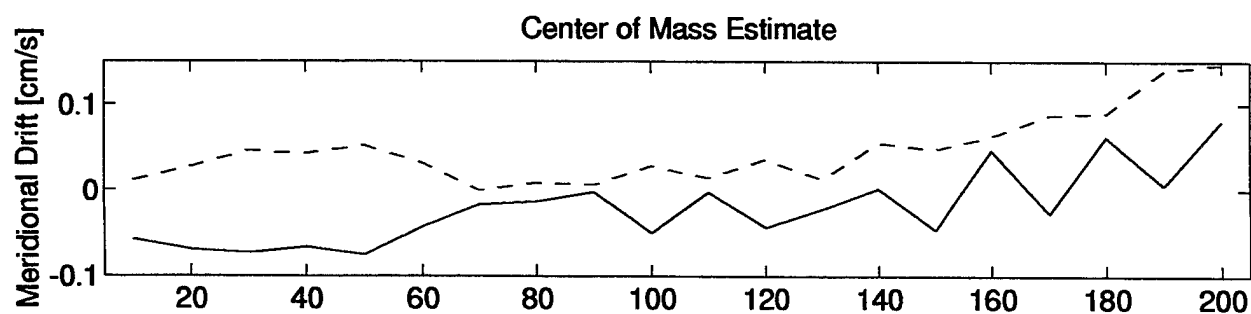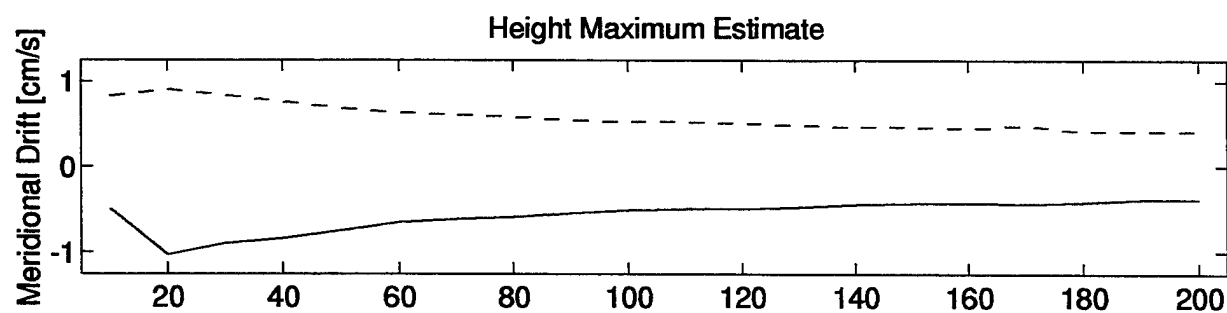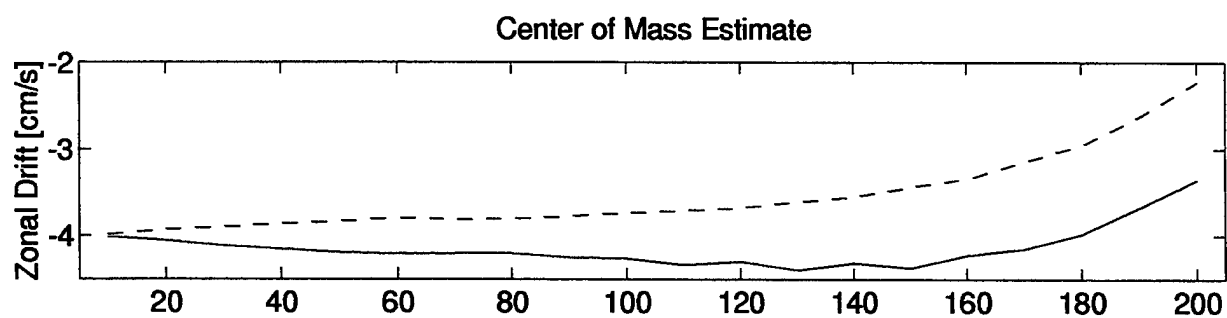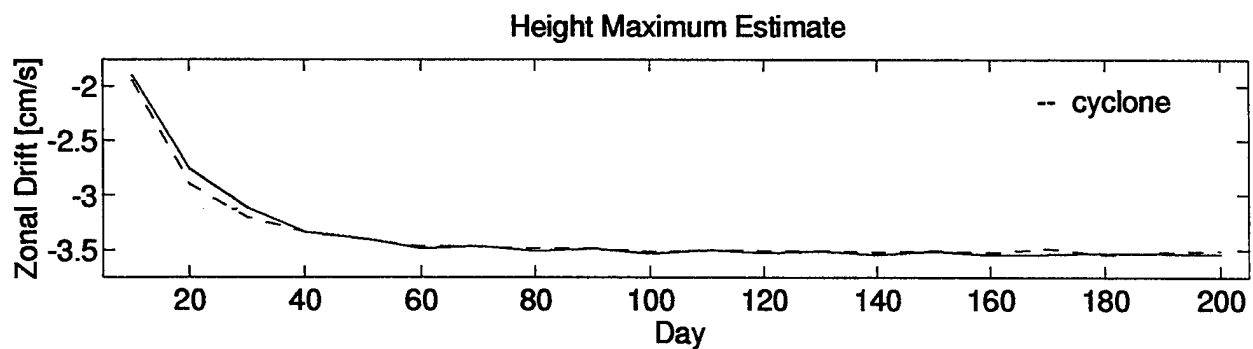
Figure 7: Zonal and meridional drift speeds for the center of mass and height extrema for a 300 m cyclone (dashed line) and anticyclone (solid line) on an f-plane with the topographic slope doubled to 2×10^{-4}.

Table 2: Summary of Experimental Results

h_0	β/f-plane	Slope	CM Zonal	CM Meridional	HM Zonal	HM Meridional
+200	f	10^{-4}	-1.79/-1.98	$1.85/2.29 \times 10^{-2}$	-1.6/-.74	.33/.41
-200	f	10^{-4}	-2.27/-2.02	$-2.82/-4.13 \times 10^{-2}$	-1.59/-.74	-.32/-.16
+300	f	10^{-4}	-1.85/-1.99	$1.74/.573 \times 10^{-2}$	-1.70/-.85	.31/.45
-300	f	10^{-4}	-2.18/-2.01	$-2.16/-3.12 \times 10^{-2}$	-1.69/-.80	-.31/-.22
+300	f	10^{-4}	-1.86/-1.99	$1.06/.556 \times 10^{-2}$	-1.70/-.95	.31/.32
-300	f -	10^{-4}	-2.17/-2.00	$-1.97/-2.74 \times 10^{-2}$	-1.71/-.95	-.31/-.32
+500	f	10^{-4}	-1.89/-1.99	$1.0/1.37 \times 10^{-2}$	-1.80/-1.14	.26/.44
-500	f	10^{-4}	-2.11/-2.0	$-1.38/-2.88 \times 10^{-2}$	-1.82/-1.11	-.27/-.41
+200	β	0	-3.39/-4.30	$4.09/-.419 \times 10^{-2}$	-3.33/-1.82	.64/.65
-200	β	0	-4.18/-3.75	$-2.26/-8.85 \times 10^{-3}$	-2.39/-1.27	-.86/-.54
+300	β	10^{-4}	-5.62/-6.09	$2.38/22.7 \times 10^{-3}$	-4.79/-3.05	1.12/.96
-300	β	10^{-4}	-5.54/-5.23	$-3.46/-5.04 \times 10{-2}$	-4.06/-2.2	-1.04/-.92
+300	f	2×10^{-4}	-3.80/-3.98	$2.49/1.19 \times 10^{-2}$	-3.29/-1.94	.65/.83
-300	f	2×10^{-4}	-4.20/-4.01	$-4.09/-5.73 \times 10^{-2}$	-3.27/-1.89	-.64/-.50
-200	f	0	$-9.17/-7.91 \times 10^{-3}$	$-9.19/-10.4 \times 10^{-3}$	-3.32×10^{-3} /.14	$-1.05/-.828 \times 10^{-3}$

324

Although baroclinic instabilities are excluded from the reduced gravity system, barotropic instabilities are possible. Smith and O'Brien (1983) do point out, however, that the Gaussian vortex is a stable feature. While the vortex does become distorted as the dispersive field develops, the radial gradient of the azimuthally-averaged potential vorticity must change sign to meet the necessary condition for instability (Carton and McWilliams, 1989). This appears an unlikely development.

(4) The (1,1) barotropic basin modes has perhaps been excited.

(5) Numerical error
Discrete sampling of the slowly-moving center may result in "jumpy" position values.

An experiment was run with a 200 m vortex on the β-plane with no topography. Figure 8 presents the results for the zonal and meridional drifts. Again, the height maximum moves faster meridionally than the center of mass. As with the doubled-slope f-plane case, the center of mass value decreases from about -4 cm s^{-1}, the theoretical value, to near $-2 - 3$ cm s^{-1}. There is a strong difference between cyclones and anticyclones. Note, too, that a reversal occurs as to which moves faster. Both centers of mass move northward after about day 150. Again, this is due to wall interaction.

Figure 9 presents analogous information for an experiment run on an f-plane with a flat bottom. Of greatest importance is that the twenty day oscillation occurs here as well. This means that the Rossby wave wake mechanism must be ruled out, as there exists in this case no way to produce a Rossby wave. This leaves the possibility of an instability, the first basin mode, or numerical error. Also of interest, but not easily explained at the moment, is that the zonal velocity of the height maximum is initially .14 cm s^{-1} and asymptotes to zero.

After Longuet-Higgins (1964), the period for a barotropic mode (m,n) in a basin of size (a × b) is

$$ T = \frac{4\pi}{\beta}(\frac{m^2\pi^2}{a^2} + \frac{n^2\pi^2}{b^2})^{1/2} $$

The first mode has a period of sixteen days, leaving it a reasonable culprit.

Finally, figures 10-11 show the result of an experiment run with a 300 m cyclone in initial gradient balance. The pressure was sampled every ten days for the first sixty days of the model run, after which it was increased to every four hours for the duration of the 150-day experiment. It is clear that there is indeed an inertial oscillation superimposed upon the topographically-induced translation. The amplitude of the oscillation is a much smaller part of the height extremum's meridional drift but does dominate the center of mass motion. While as stated above no physical mechanism exists to excite an inertial oscillation, there is a strong tendency for numerical models to accumulate energy at the inertial frequency, which is a free mode in the system. Also of interest is that within the height extremum estimates there appears to be a longer-term oscillation, again of twenty to twenty-five days, upon which the inertial oscillation is superimposed.

Simple spectral analysis of these densely-sampled self-advection speeds is shown in figures 12-13. Briefly, an average value was removed from all records except the zonal center of mass, from which a linear trend was removed. The records were then split into two pieces and the power density estimates at each frequency averaged in order to improve confidence levels. The basin mode frequency (16 day period) and the inertial frequency (17 hour period) are marked by circles. Results are somewhat inconclusive. The center of mass speeds do show higher power at the lower frequency, although this is not the case in the displacement maximum speeds. The 95% confidence level, not shown, would indicate that the basin mode frequency does indeed contain higher power; however, this interpretation is complicated by the fact that the basin mode frequency is the fourth spectral estimate and thus its validity somewhat suspect. It must also be acknowledged that in this case frequency averaging would have been preferable to piece-averaging, given the previously observed changes in drift speed accompanying vortex evolution.

4 Corrections to the Center of Mass Translation Expression

Two second order corrections to the center of mass translation are examined to try to reduce the formal error and in hopes of explaining the small but finite meridional drift seen in all the experimental
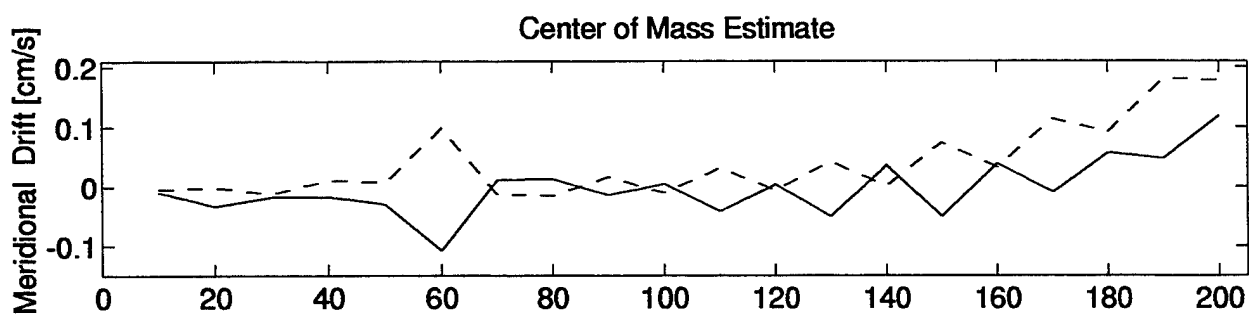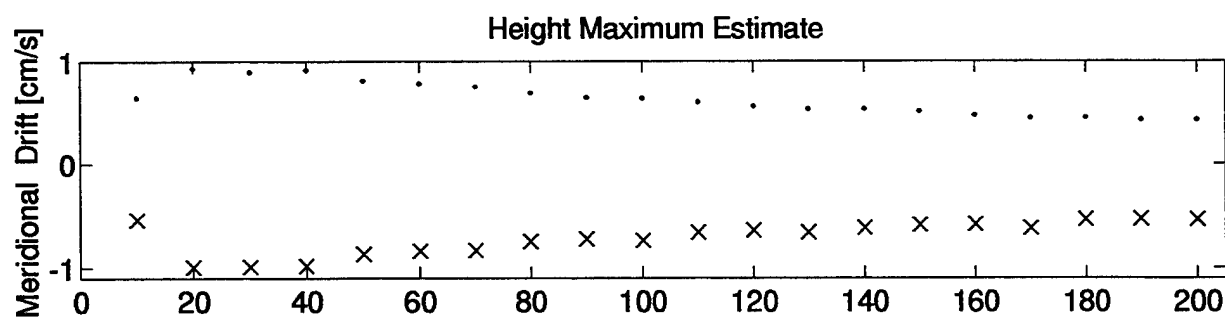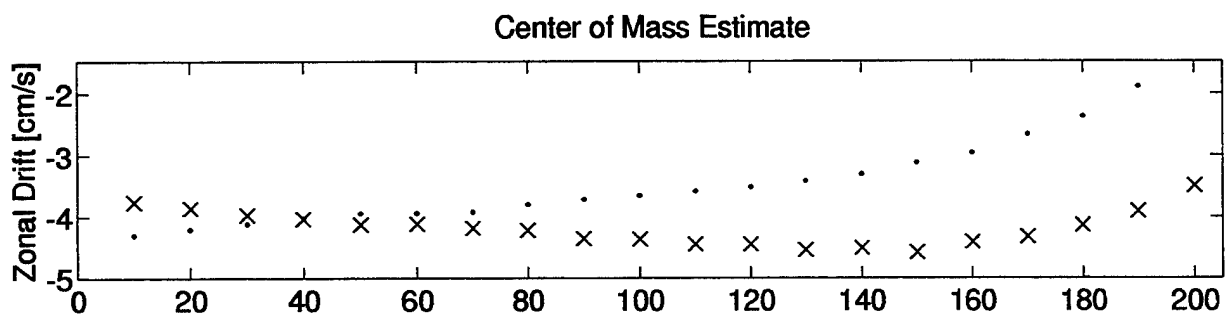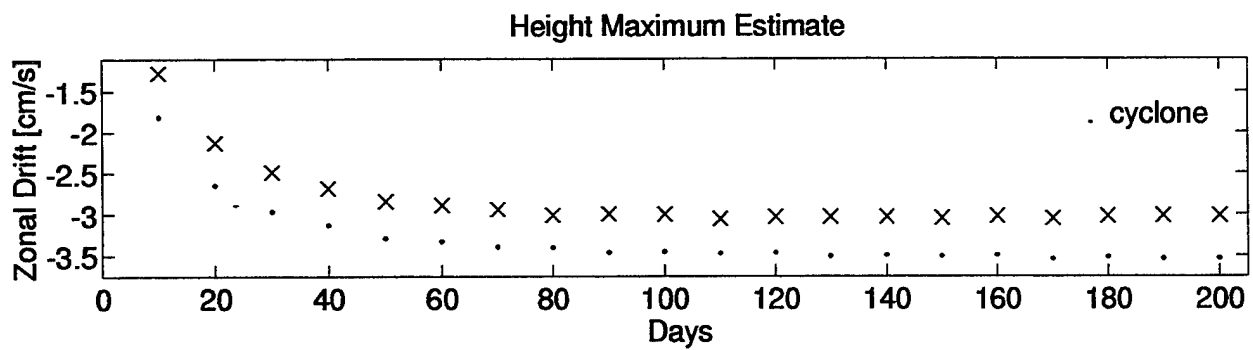
325

Figure 8: Zonal and meridional drift speeds for the center of mass and height extrema for a 200 m cyclone (dashed line) and anti-cyclone (solid line) on a β-plane with a flat bottom.
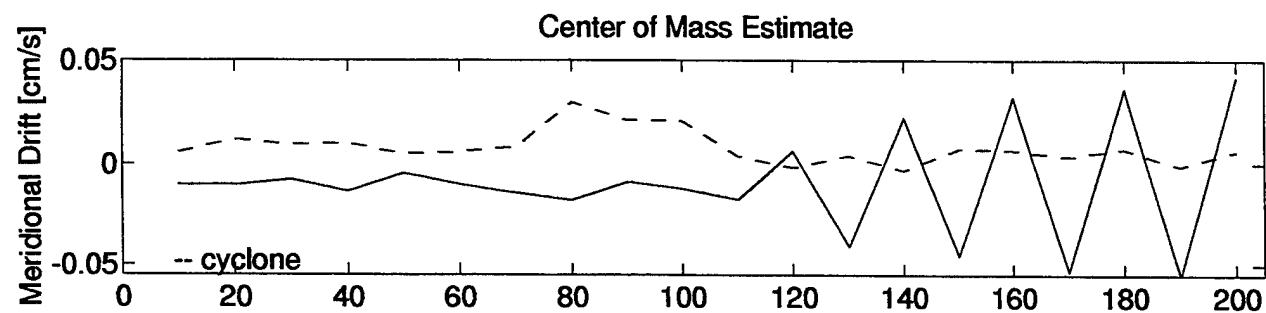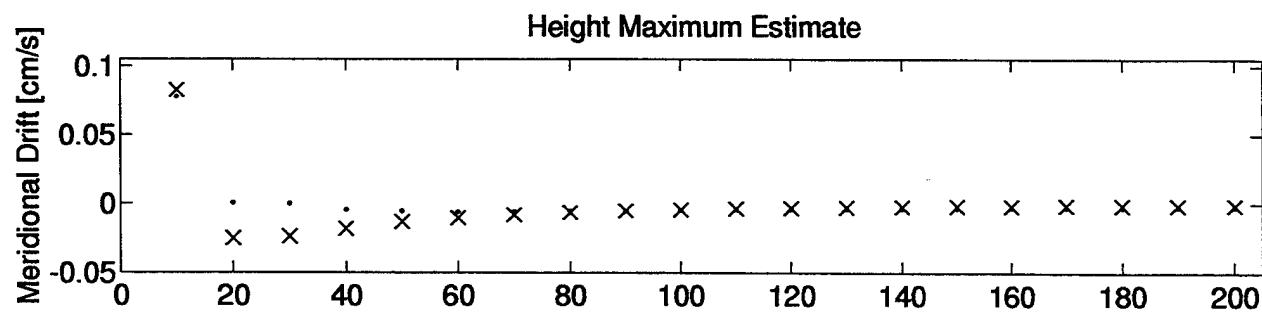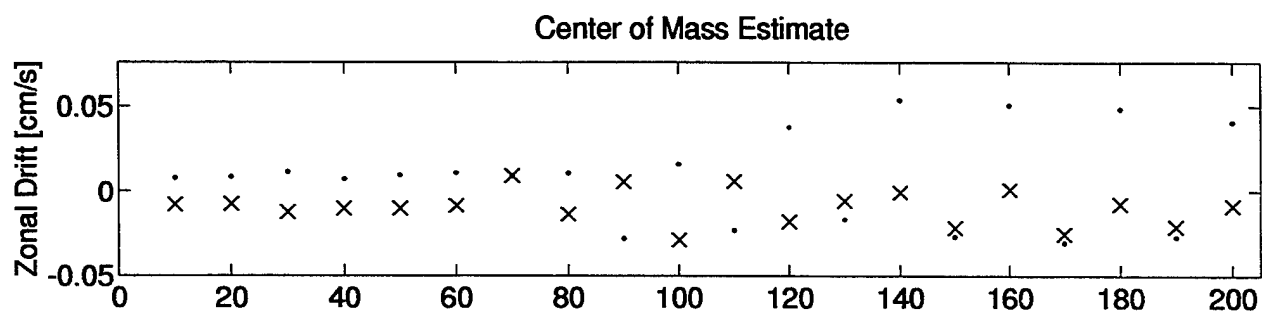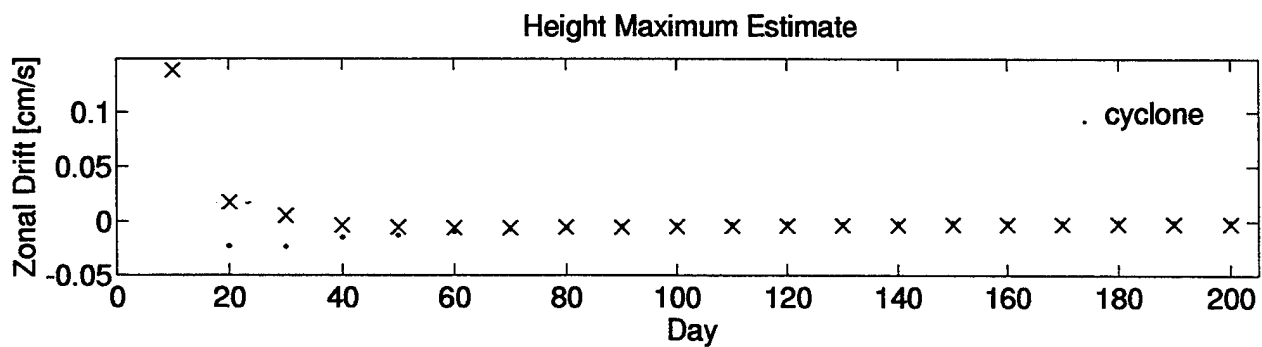
Figure 9: As in Figure 8, but for vortices on an f-plane with a flat bottom.
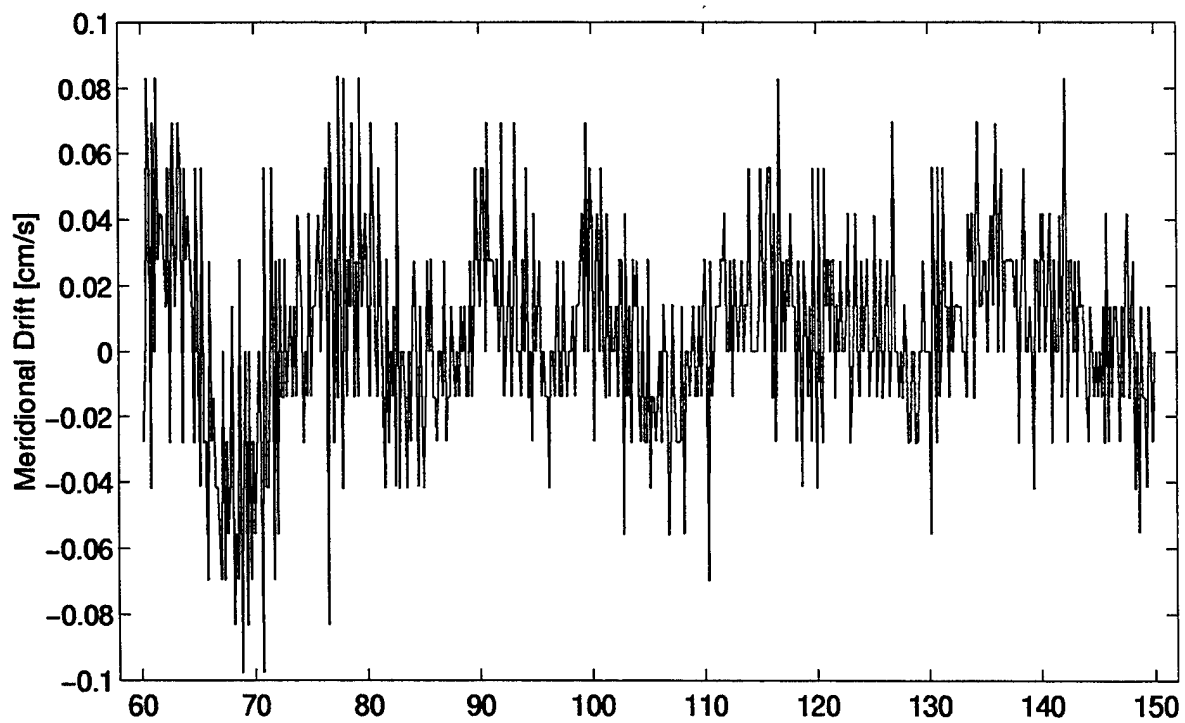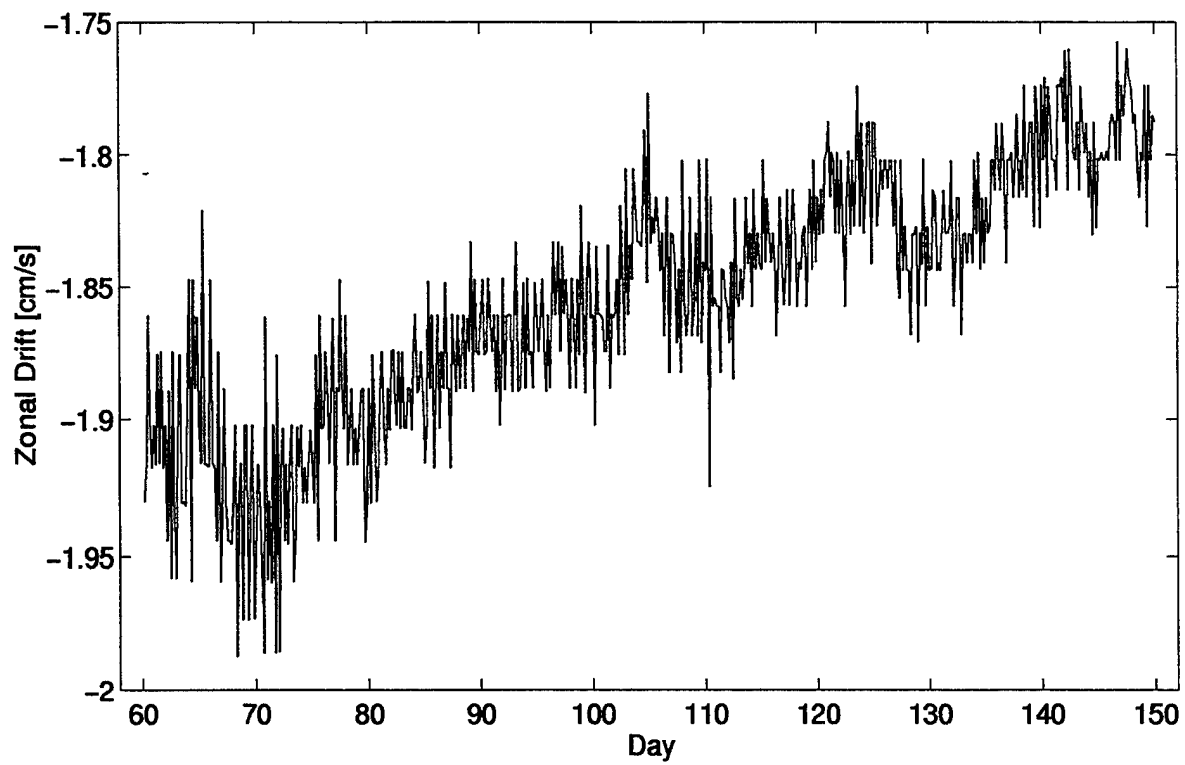
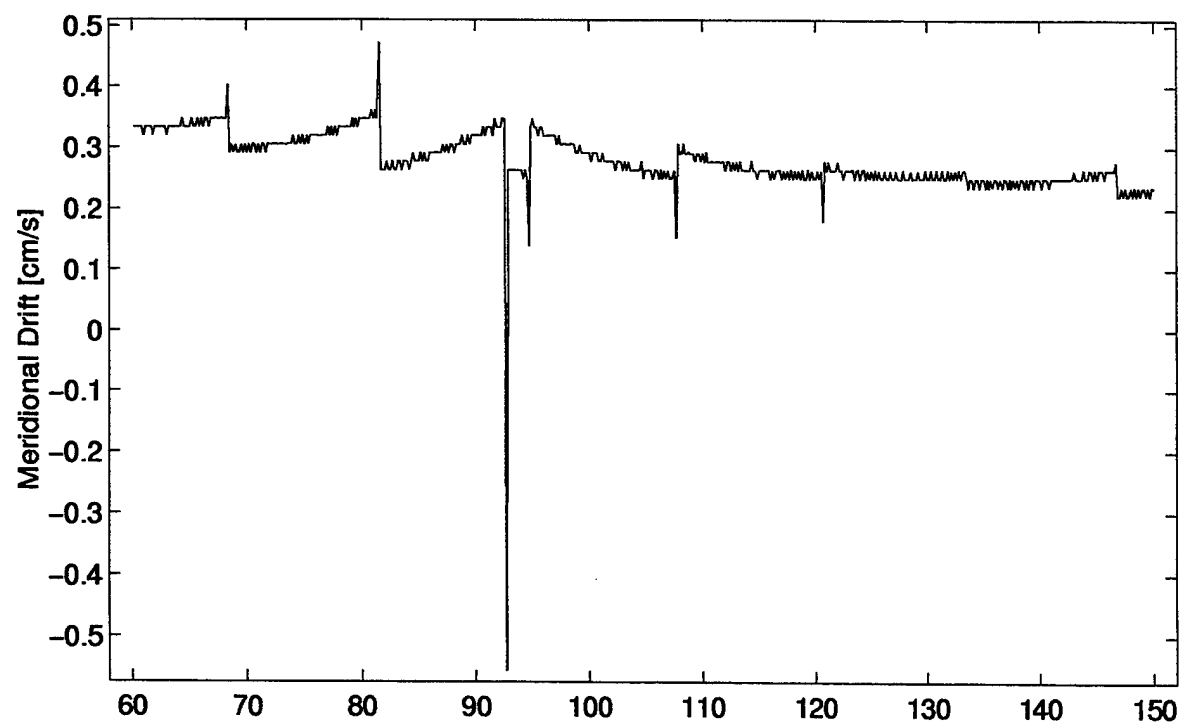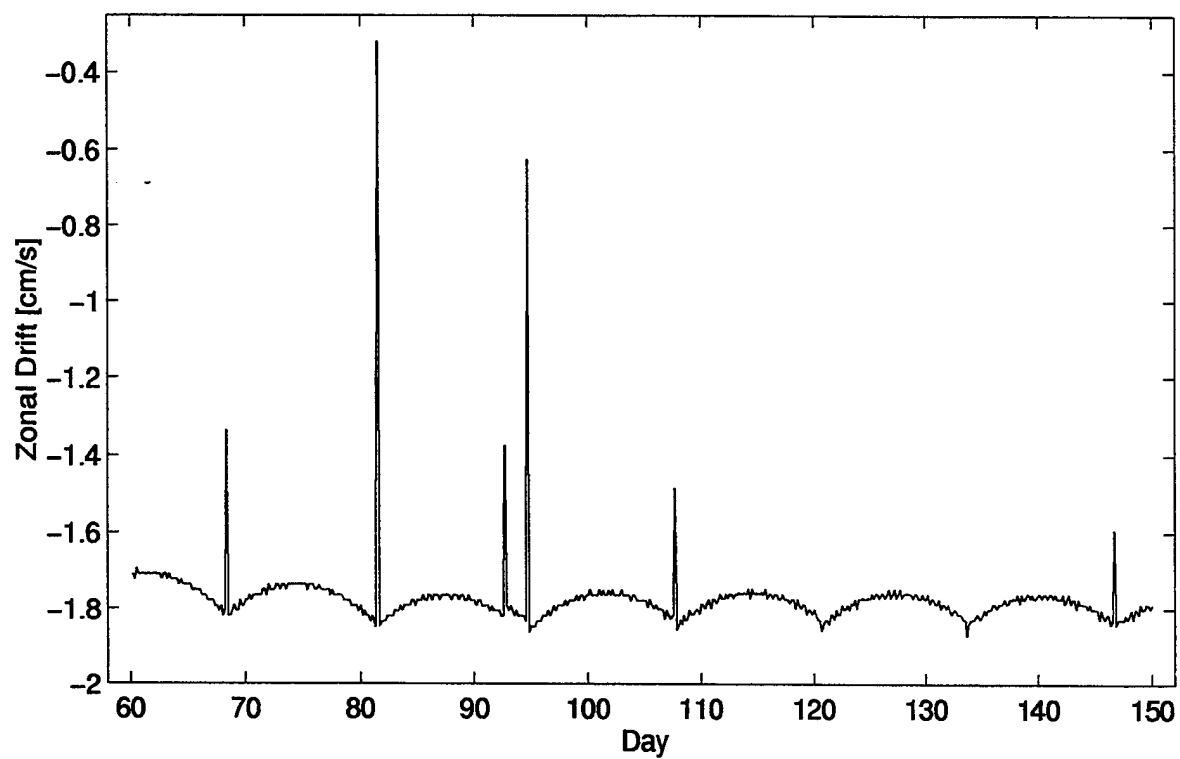Figure 10: Center of mass drift speeds for 150-day experiment sampled every 4 hours.

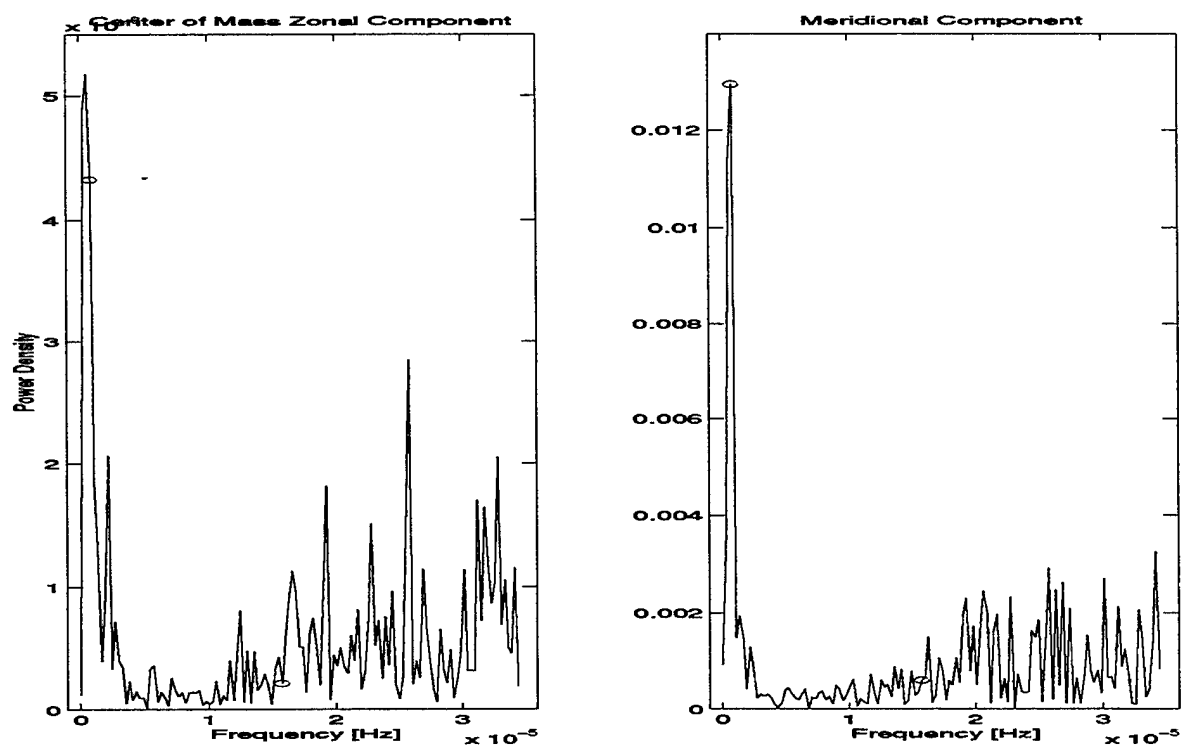Figure 11: Height extremum drift speeds for 150-day experiment sampled every 4 hours.

329

Figure 12: Power spectrum of center of mass drift speeds from 150-day experiment sampled every 4 hours.
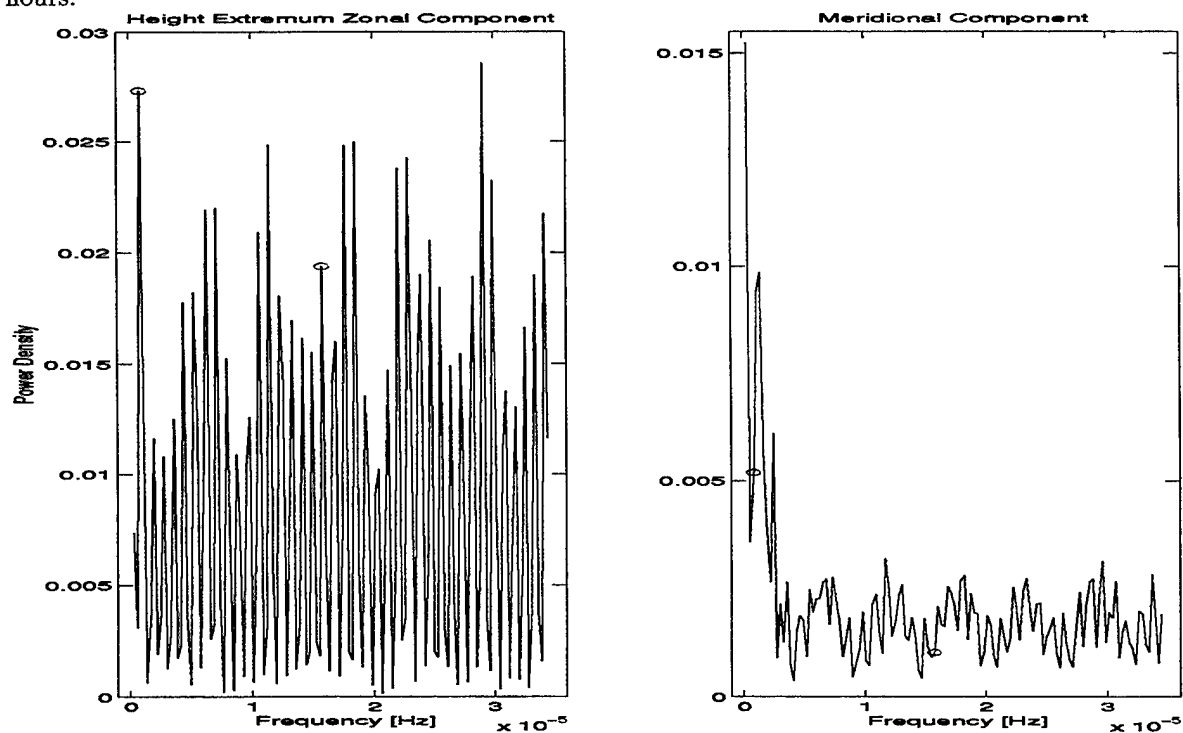

Figure 13: Power spectrum of height extremum drift speeds from 150-day experiment sampled every 4 hours.

runs, as well as the circulation–dependent asymmetry of the zonal drift speed about the theoretical value. These corrections were selected in an attempt to isolate distinct effects; the first evaluates the second order effect in the geostrophic calculation, and the second attempts to assess the leading order ageostrophic effect. This is assumed to be the centripetal acceleration term of gradient balance flow.

A   Second Order Geostrophic Corrections

<u>f-Plane</u>:

The second order correction to the geostrophic velocity approximation is derived following Chassignet (1989).

Take the first order equations

$$\frac{d^2X}{dt^2} - f_o\frac{dY}{dt} = -g'\langle b_x \eta\rangle \tag{5}$$

$$\frac{d^2Y}{dt^2} + f_o\frac{dX}{dt} = -g'\langle b_y \eta\rangle \tag{6}$$

and differentiate with respect to time:

$$\frac{d^3X}{dt^3} - f_o\frac{d^2Y}{dt^2} = -g'\langle b_x \eta_t\rangle \tag{7}$$

$$\frac{d^3Y}{dt^3} + f_o\frac{d^2X}{dt^2} = -g'\langle b_y \eta_t\rangle \tag{8}$$

Now, substitute the expression for $\frac{d^2Y}{dt^2}$ from (7) into (6) and that for $\frac{d^2X}{dt^2}$ in (8) into (5).

$$\frac{d^3X}{dt^3} + f_o^2\frac{dX}{dt} = g'(-f_o\langle b_y\eta\rangle + \langle b_x(uh)_x\rangle + \langle b_x(vh)_y\rangle$$

$$\frac{d^3Y}{dt^3} + f_o^2\frac{dY}{dt} = g'(f_o\langle b_x\eta\rangle + \langle b_y(uh)_x\rangle + \langle b_y(vh)_y\rangle$$

Now, if ω is small enough to warrant dropping $\frac{d^2X}{dt^2}$ and $\frac{d^2Y}{dt^2}$, then the third derivatives, $O(\omega^2)$ should be dropped, leaving the first order equations. Thus, for constant slope, the f-plane equations do not change at higher order.

II. β-Plane

$$\frac{d^2X}{dt^2} - f_o\frac{dY}{dt} = -g'\langle b_x\eta\rangle + \beta\langle hyv\rangle \tag{9}$$

$$\frac{d^2Y}{dt^2} + f_o\frac{dX}{dt} = -g'\langle b_y\eta\rangle + \beta\langle hyu\rangle \tag{10}$$

Similar operations yield

$$\frac{d^3X}{dt^3} + f_o^2\frac{dX}{dt} =$$

$$-g'f_o\langle b_y\eta\rangle - 2\beta f_o\langle hyu\rangle - \beta^2\langle y^2 hu\rangle + g'[-\beta\langle yh\eta_y\rangle + \langle b_x(uh)_x\rangle + \langle b_x(vh)_y\rangle]$$

331

$$\frac{d^3Y}{dt^3} + f_o^2 \frac{dY}{dt} =$$

$$g'f_o \langle b_x \eta \rangle - 2\beta f_o \langle hyv \rangle - \beta^2 \langle y^2 hv \rangle + g'[\langle b_y(uh)_x \rangle + \langle b_y(vh)_y \rangle - \beta g' \langle by \eta_x \rangle] - \beta \langle huv \rangle$$

Again assuming constant slope and negligible third derivatives, these expressions can be compared to those of Chassignet (1989), derived in exactly the same manner; new terms due to topography are:

$$\frac{dX}{dt} = -\frac{g'b_y}{f_o} + \frac{\beta}{f_o}[2\langle byu \rangle - \frac{\beta}{f_o}\langle y^2 bu \rangle + \frac{g'}{f_o}\langle yb\eta_y \rangle]$$

$$\frac{dY}{dt} = -\frac{g'b_x}{f_o} + \frac{\beta}{f_o}[2\langle byv \rangle + \frac{\beta}{f_o}\langle y^2 bv \rangle - \frac{g'}{f_o}\langle yb\eta_x \rangle + \frac{1}{f_o}\langle buv \rangle]$$

Now, for a constant slope in y and axisymmetric geometry, integrating in r and θ gives zero meridional motion once again. Several terms are left in the zonal equation, but the largest is $O(10^{-5} \text{ cm s}^{-1})$. Thus, the second order geostrophic terms do not contribute significantly over an infinite domain. This is in keeping with the findings of Chassignet (1989) that the second order terms are important over a finite domain (within the vortex) but disappear as the distance from the center becomes large.

B   Gradient Balance Terms

The centripetal acceleration of the gradient balance is expected to be the leading order ageostrophic term. The total velocity can be written $V_{total} = V_g(1 + \alpha)$, where V_g is the geostrophic velocity and α the correction to it due to gradient balance terms.

Note that α will add to cyclones and subtract from anticyclones:

$$V = \frac{g'\eta_r}{f} + \frac{(g'\eta_r)^2}{rf^3} \qquad\qquad V = -\frac{g'\eta_r}{f} + \frac{(g'\eta_r)^2}{rf^3}$$

$$\eta_r < 0 \qquad\qquad\qquad\qquad \eta_r > 0$$

The first term gives the geostrophic velocity, while the second, higher order term gives the centripetal acceleration. A straightforward way to evaluate its contribution is to integrate the centripetal term over all space, thus obtaining an average correction:

$$\alpha = \frac{1}{V_g} \iint \frac{(g'\eta_r)^2}{rf_o^3} \, dr \, d\theta$$

This gives

$$V_{total} = V_g + \iint \frac{(g'\eta_r)^2}{rf_o^3} \, dr \, d\theta$$

$$\eta = h_o e^{\frac{-r^2}{2L^2}} \qquad\qquad \eta_r = -\frac{r}{L^2} h_o e^{\frac{-r^2}{2L^2}} = -\frac{r}{L^2}\eta$$

$$\frac{\int \frac{g'^2}{f_o^3 L^4} \int r e^{\frac{-r^2}{L^2}} \, dr \, d\theta}{\iint \frac{g'}{f_o}\eta_r \, dr \, d\theta} \qquad = \quad \frac{\frac{\pi g'^2}{f_o^3 L^2}}{\frac{g'\pi}{f_o}} \qquad = \ 0.08 \ = \ \alpha$$

The revised estimate of the center of mass translation is 6.48 ± 0.45 cm s^{-1}. Unfortunately, this does not account for the nonsymmetric split around the predicted value found between cyclones and anticycles seen in the numerical f-plane drift speeds. Also, due to the choice to integrate in r and θ rather than x and y, the finite meridional speed cannot be addressed. There is a clear need to follow the initial theory development by substituting a gradient balance approximation of the velocity

332

components into the center of mass equations and integrating in Cartesian coordinates if the meridional speed is to be calculated. This integration is fairly cumbersome and due to time constraints was left for future work.

5 Conclusions

It has been found in this study, as in others, that the center of mass is a good representation of the eddy center in terms of topographically-induced zonal motion, but that the meridional motion of the vortex center as defined by a height extremum significantly differs from that of the center of mass. The height extremum's meridionally motion is in general an order of magnitude faster than the center of mass'.

An important finding was that the f-plane case recaptured the result of Nof (1983a) that the translation was given by the expression $-g\frac{b_x}{f_o}$. Numerical simulation gives a fair match to the analytical prediction of a drift speed independent of the eddy intensity (maximum amplitude at the center). Initial speeds match it quite well, and later evolution is hypothesized to be driven by interaction with the Rossby wave wake. While an experiment with a doubled slope gives average zonal drift speeds slightly lower than predicted (about five percent), this is still well within the error estimate on the zonal speed, and again, initial speeds match the analytical prediction much better.

An oscillation with a period of approximately twenty days appears in the numerically simulated drift speeds. It is as yet unclear whether this is truly an aliased inertial oscillation or the first barotropic basin mode, or simply numerical error related to discrete sampling on a finite grid.

Future Work

There exists a rich field for future work. Clearly, ageostrophic effects, mainly the centripetal term, should be reconsidered in a more detailed and careful manner, and the nonlinear terms could be included. An asymptotic expansion of the dispersion field has been suggested as an additional way of including non-axisymmetric effects in the analytical investigation.

One of the most intriguing questions remains the result of rotating the topographic and the planetary β effects out of phase. Variations in the topography are needed as well: seamounts, ridges, shelves, and basins are all common within the ocean and easily investigated in the context of a numerical model.

Acknowledgements

References

Bleck, R. and D.B. Boudra. (1986). Wind-driven spin up in eddy-resolving ocean models formulated in isopycnic and isobaric coordinates. *J. Phys. Oceanogr.*, 91, 7611-7621.

Carton, X.J. and J.C. McWilliams (1989). Barotropic and baroclinic instabilities of axisymmetric vortices in a quasigeostrophic model. *In*: Mesoscale/synoptic coherent structures in geophysical turbulence. J.C.J. Nihoul and B.M. Jamart, Eds. Elsevier, 841 pp.

Chassignet, E.P. (1989). On the meridional propagation of isolated eddies. WHOI Tech. No. WHOI-89-54, 236-259.

Chassignet, E.P. and B. Cushman-Roisin (1991). On the influence of a lower layer on the propagation of nonlinear oceanic eddies. *J. Phys. Oceanogr.*, 21, 939-957.

Chassignet, E.P., D.B. Olson, and D.B. Boudra. (1990). Motion and evolution of oceanic rings in a numerical model and in observations. *J. Geophys. Res.*, **95**, 22121–22140.

Cheney, R.E. and P.L. Richardson (1976). Observed decay of a Gulf Stream ring. *Deep-Sea Res.*, **23**, 143-155.

Cushman-Roisin, B., E.P. Chassignet, and B. Tang. (1990). Westward motion of mesoscale eddies. *J. Phys. Ocean.*, **20**, 758–768.

Longuet-Higgins, M.S. (1964). Planetary waves on a rotating sphere. *Proc. Roy. Soc.*, **279**, 446-473.

McWilliams, J. and G. Flierl. (1979). On the evolution of isolated, nonlinear vortices, with applications to Gulf Stream rings. *J. Phys. Ocean.*, **14**, 1155-1192.

Mied, R.P. and G.J. Lindemann. (1979). The propagation and evolution of cyclonic Gulf Stream rings. *J. Phys. Oceanogr.*, **9**, 1193-1206.

Nof, D. (1981). On the β-induced movement of isolated baroclinic eddies. *J. Phys. Oceanogr.*, **11**, 1162-1672.

Nof, D. (1983a). The translation of isolated cold eddies on a sloping bottom. *Deep-Sea Res.*, **30**, 171–182.

Nof, D. (1983b). On the migration of isolated eddies with application to Gulf Stream rings. *J. Mar. Res.*, **41**, 399-425.

Olson, D.B. (1991). Rings in the Ocean. *Annu. Rev. Earth Planet. Sci.* , **19**, 283–311.

Olson, D.B. and R.H. Evans (1986). Rings of the Agulhas Current. *Deep- Sea Res.*, **33**, 27–42.

Smith, D.C., IV, and J.J. O'Brien (1983). The interaction of a two-layer isolated mesoscale eddy with bottom topography. *J. Phys. Oceanogr.*, **13**, 1681-1697.

Toward an Improved Dispersion Operator
For Organisms With Long Mean Free Paths

Douglas T. Morgan

Thayer School of Engineering, Dartmouth College

Hanover, New Hampshire 03755

September 16, 1994

1   Introduction

The typical approximation that a population of organisms disperses through random movements according to standard Fickian diffusion is often fine as long as the mean free path of an organism is small as compared to the scale of spatial variations in the population. If, however, the mean free path of the organism is much larger than the scale of variation of the population, this approximation is no longer strictly valid. In an effort to account for organisms with long mean free paths, we borrow from radiative transfer work (Stein and Spiegel, 1967, Unno and Spiegel, 1966) to derive a new dispersion operator and test its effects on a simple one dimensional problem.

In modelling a spatially varying population of organisms, we often keep track of just the number of individuals at a given point. If we call this number of individuals n, where n is a function of space x and time t, the rate that this number n changes with time can be modelled as a combination of two functions, one function, $f(n; x, t)$, representing the population dynamics (birth and death) at point x, and one function, $g(n; x, t)$, representing the dispersal of organisms to and from point x.

$$\partial_t n = f(n; x, t) + g(n; x, t) \tag{1}$$

The function f can contain a variety of interesting population dynamics. Later, this study investigates one of the simplest, logistic growth, but for now this study concentrates on the dispersal function g.
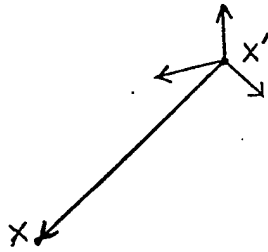
335

Figure 1: Organism at point x' moves to point x

2   The Integral Basis of the Dispersion Operator

The dispersal function g which describes how individuals in a population spread from point to point is most often modelled with Fickian diffusion as expressed by the Laplacian operator. This, however, is an approximation of a more general integral operator, and may not always be the best approximation for a the way a given population spreads spatially. Working from basic principles we derive the integral dispersion operator and then three approximations to it: the regular Laplacian diffusion, a fourth order series truncation, and a new rational approximation.

Given two points in space x and x', let $K(x, x')$ be the chance that an organism starting at point x' at one instant will be at point x the next instant as illustrated in Figure (1). Although $K(x, x')$ could take many forms to model many different behavior patterns, $K(x, x')$ will commonly be some "peaky" function depending on the separation of points x and x' where there is a high probability of an organism moving between points a short distance apart and a sharply decreasing probability of an organism moving between points longer distances apart. A Gaussian distribution is a good example of this type of curve.

For our purposes we assume that, as sketched in Figure (2), this function $K(x, x')$ is symmetric such that it depends only the absolute separation distance, that the spatial variations in $K(x, x')$ can be described by a length scale l, and that $K(x, x')$ has been normalized to have unit integral with respect to x.

Organisms already at point x disperse as a function of the number of individuals at that point. Thus, similar to radiative heat transfer (Stein and Spiegel, 1967), neglecting all retardation effects from travel time, the dispersal of organisms to and from point x is given by multiplying $K(x - x')$ by the number of individuals at each point x', and integrating over all possible points x' to find the number of individuals who move to point x, and then subtracting the number of individuals already at point x who choose to leave;
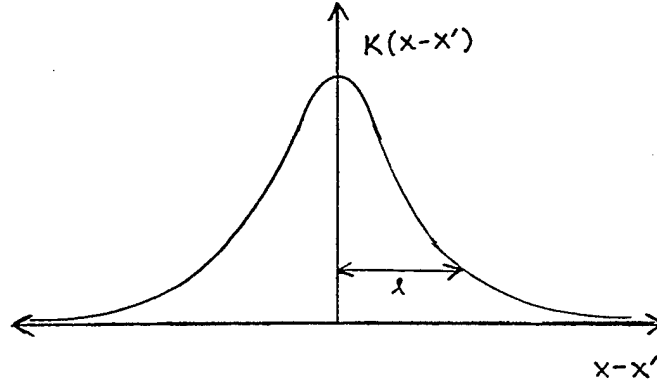
Figure 2: An example of a function $K(x, x')$, the probability of moving from point x' to point x.

$$g(x) = \int_{-\infty}^{\infty} aK(x - x')n(x', t)dx' - an(x, t) \qquad (2)$$

where a is a global "activity" constant relating the tendency of individuals to move from a point.

We can define a new spatial variable:

$$\tilde{x} = x' - x \qquad (3)$$

simplify our dispersion kernel:

$$\hat{K}(x - x') = K(x - x') - \delta(x - x') = \hat{K}(\tilde{x}) \qquad (4)$$

and rewrite Eq. (2)

$$g(x) = \int_{-\infty}^{\infty} a\hat{K}(\tilde{x})n(\tilde{x} + x, t)d\tilde{x} \qquad (5)$$

Eq. (5) is the exact integral dispersion operator. Given our assumptions on $K(x, x')$, however, we can derive several approximations to this exact operator that are more efficient to deal with in most circumstances. If n has a length scale L that is much larger than the length scale l of $\hat{K}(\tilde{x})$, we can Taylor expand $n(\tilde{x} + x, t)$ around $n(x, t)$

$$n(\tilde{x} + x, t) = n(x, t) + \tilde{x}\partial_x n + \frac{1}{2}\tilde{x}^2 \partial_x^2 n + \frac{1}{6}\tilde{x}^3 \partial_x^3 n + \frac{1}{24}\tilde{x}^4 \partial_x^4 n + ... \qquad (6)$$

Substituting this into Eq. (5)

$$g(x) = \int_{-\infty}^{\infty} a\hat{K}(\tilde{x})[n(x,t) + \tilde{x}\partial_x n + \frac{1}{2}\tilde{x}^2\partial_x^2 n + \frac{1}{6}\tilde{x}^3\partial_x^3 n + \frac{1}{24}\tilde{x}^4\partial_x^4 n + ...]d\tilde{x} \quad (7)$$

$$g(x) = an\int_{-\infty}^{\infty} \hat{K}(\tilde{x})d\tilde{x} + a\partial_x n\int_{-\infty}^{\infty} \tilde{x}\hat{K}(\tilde{x})d\tilde{x} + a\partial_x^2 n\int_{-\infty}^{\infty}\frac{1}{2}\tilde{x}^2\hat{K}(\tilde{x})d\tilde{x} + ... \quad (8)$$

If we require the integral of $\hat{K}(\tilde{x})d\tilde{x} = 0$ then the first term on the RHS of Eq. (8) integrates to zero. The second term, and likewise all other terms with odd powers of $\tilde{x}$, integrates to zero because of the symmetry of $\hat{K}(\tilde{x})$. This leaves only the RHS terms with even powers of $\tilde{x}$ in the integrals.

$$g(x) = a\partial_x^2 n\int_{-\infty}^{\infty}\frac{1}{2}\tilde{x}^2\hat{K}(\tilde{x})d\tilde{x} + a\partial_x^4 n\int_{-\infty}^{\infty}\frac{1}{24}\tilde{x}^4\hat{K}(\tilde{x})d\tilde{x} + ... \quad (9)$$

$$g(x) = D\partial_x^2 n + F\partial_x^4 n + H\partial_x^6 n + ... \quad (10)$$

where D, F, and H are constants, with $D > F > H$. Truncating this series after the first term gives our first approximation, the normal diffusion operator:

$$g(x) \approx D\partial_x^2 n \quad (11)$$

Truncating series Eq. (10) at fourth order gives our second approximation:

$$g(x) \approx D\partial_x^2 n + F\partial_x^4 n \quad (12)$$

Rewritting series Eq. (10) as

$$g(x) = D\partial_x^2[1 + \frac{F}{D}\partial_x^2 + \frac{H}{D}\partial_x^4 + ...]n \quad (13)$$

and making use of the Pade rational approximation, gives our third approximation:

$$g(x) \approx D\frac{\partial_x^2}{1 - \frac{F}{D}\partial_x^2}n \quad (14)$$

This is our desired new dispersion operator. Generalizing our operator Eq. (14) to 2D or 3D gives:

$$g(x) \approx D\frac{\nabla^2}{1 - \alpha\nabla^2}n \quad (15)$$

3 Comparison of Approximate Dispersion Operators

To examine the properties of our new operator as compared to the other approximations and the exact integral operator, look at solutions of $\partial_t n = g(n; x, t)$ with $n(x, t) = n_o e^{ikx} e^{-\sigma t}$, where k is the spatial wave number and σ is the real decay rate.

- For regular diffusion

$$g(x) = \partial_t n \approx D \nabla^2 n \tag{16}$$

$$\sigma \approx D k^2 \tag{17}$$

- For a normal series truncation of Eq. (10) at fourth order

$$g(x) = \partial_t n \approx D \nabla^2 n + F \nabla^4 n \tag{18}$$

$$\sigma \approx D k^2 - F k^4 \tag{19}$$

- For the new approximate dispersion operator

$$g(x) = \partial_t n \approx D \frac{\nabla^2}{1 - \alpha \nabla^2} n \tag{20}$$

$$\sigma \approx D \frac{k^2}{1 + \alpha k^2} \tag{21}$$

- For the exact expression Eq. (5)

$$g(x) = \partial_t n = \int_\infty^\infty a \hat{K}(\tilde{x}) n_o e^{ik(\tilde{x}+x)} e^{\sigma t} d\tilde{x} \tag{22}$$

$$\partial_t n = a n_o e^{ikx} e^{\sigma t} \int_\infty^\infty \hat{K}(\tilde{x}) e^{ik\tilde{x}} d\tilde{x} = a n_o e^{ikx} e^{\sigma t} \mathrm{T}(k) \tag{23}$$

where $\mathrm{T}(k)$ is the Fourier Transform of $\tilde{K}(\tilde{x})$. For $K(\tilde{x}) =$ a Gaussian distribution:

$$\tilde{K}(\tilde{x}) = \frac{1}{l\sqrt{\pi}} e^{-\frac{\tilde{x}^2}{l^2}} - \delta(\tilde{x}) \tag{24}$$

$$\mathrm{T}(k) = e^{-k^2 4 l^2} - 1 \tag{25}$$

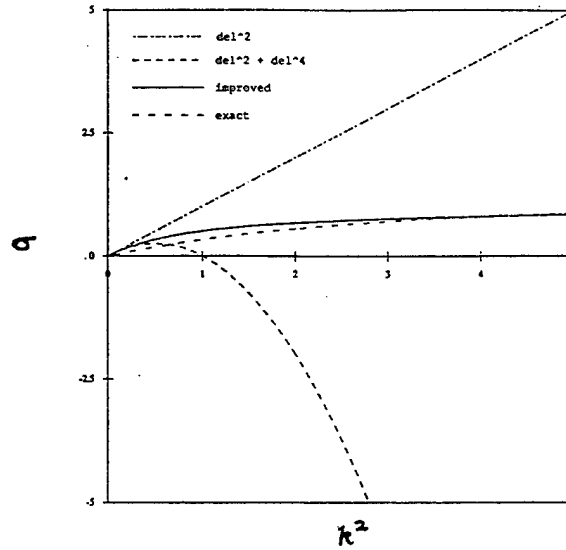$$\sigma = a(1 - e^{-k^2 4 l^2}) \tag{26}$$

339

Figure 3: Decay rate σ vs. spatial wave number k^2 for different dispersion approximations.

Figure (3) shows a plot of σ vs. k^2 for the regular diffusion operator, the truncated at fourth order operator, the new approximate operator and for the exact expression given a Gaussian kernel. Note that while the decay rates for the regular diffusion operator and the truncated series approximation both are unbounded as $k^2 \to \infty$, those for the improved operator and the exact equation go asymptotically to constants. For large k^2:

- For the new approximate operator

$$\sigma \approx \frac{D}{\alpha} \tag{27}$$

- For the exact expression

$$\sigma \approx a \tag{28}$$

By choosing $\frac{D}{\alpha} = a$ the decay rate for the new approximate operator can be made to approach the same constant as that of the exact equation for large k^2.

4 Effects of the New Dispersion Operator on a Simple Population Model

The previous section ignored the presence of dynamics such as growth and death which maybe affecting the population in order to derive a new dispersion operator. In this section we will investigate the effects of the new operator on a population model incorporating the dynamics of growth and death. These dynamics are contained in
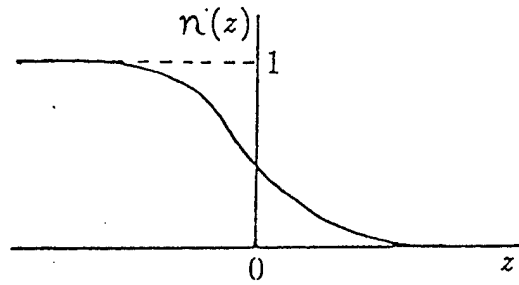
Figure 4: Traveling wave front solution to the Fisher Equation, from Murray (1989).

the function f of Eq. (1). One of the simplest dynamics that leads to interesting behavior is logistic growth. Incorporating logistic growth into function f of Eq. (1) and using the usual diffusion operator for function g gives an expression known as the Fisher Equation:

$$\partial_t n = n(1 - n) + \partial_x^2 n \tag{29}$$

Using logistic growth for f and our new dispersion operator for g, we can define an expression analogous to the Fisher equation for our new operator:

$$\partial_t p = p(1 - p) + D\frac{\partial_x^2}{1 - \alpha\partial_x^2}p \tag{30}$$

The Fisher Equation is relatively simple and well studied (see Murray, 1989, Mollison, 1977) but is interesting because it is the prototype equation which exhibits traveling wave solutions of the form

$$n(x, t) = n(z), \qquad z = x - ct \tag{31}$$

as shown in Figure (4).

For solutions of this type

$$\frac{\partial}{\partial t} = \frac{\partial z}{\partial t}\frac{\partial}{\partial z} = -c\frac{d}{dz} \tag{32}$$

• For the Fisher equation, the traveling wave solution gives

$$-cn' = n(1 - n) + n'' \tag{33}$$

341

- For the analogous expression with the new operator, the traveling wave solution gives

$$- cp' - p(1 - p) + \alpha cp''' - \alpha \left[2p'^2 + 2pp'' - p'' \right] = p'' \tag{34}$$

While the traveling wave solution to the Fisher equation gives a *second* order nonlinear ordinary differential equation, the traveling wave solution for the analogous expression gives a *third* order nonlinear ordinary differential equation, a much more complicated equation to solve.

For the second order Fisher equation we can define an equivalent system of *two* first order ODE's. With $U = n(z)$,

$$U' = V \tag{35}$$
$$V' = -cU' - U(1 - U) \tag{36}$$

This system of two first order ODE's has two fixed points, one at $(0,0)$, and the other at $(1,0)$. Linear stability analysis about these fixed points gives

- at $(1,0)$:

$$\lambda = \frac{1}{2} \left[-c \pm \sqrt{c^2 + 4} \right] \tag{37}$$

One λ has negative real part while the other has positive real part, making the fixed point $(1,0)$ a saddle point.

- at $(0,0)$:

$$\lambda = \frac{1}{2} \left[-c \pm \sqrt{c^2 - 4} \right] \tag{38}$$

Both λ have negative real parts for all positive values of c. For $c^2 \geq 4$ both λ have zero imaginary parts, making the fixed point $(0,0)$ a stable node. For $c^2 < 4$ the λ are complex conjugate pairs with non-zero imaginary part, making the fixed point $(0,0)$ a stable spiral, see Figure (5). Because a trajectory spiraling into $(0,0)$ would lead to negative values of $n(z)$, as a population model the Fisher equation requires $c \geq 2$ for traveling wave front solutions.

The traveling wave front solution (for $c > 0$) is expressed by the conditions

$$\lim_{z \to \infty} n(z) = 0 \quad \text{and} \tag{39}$$
$$\lim_{z \to -\infty} n(z) = 1 \tag{40}$$

This can be visualized in phase space as trying to connect the unstable manifold leaving point $(1,0)$ to one of the stable manifolds entering point $(0,0)$ see Figure (6). When the full PDE for Fishers equation is integrated from initial step conditions, the
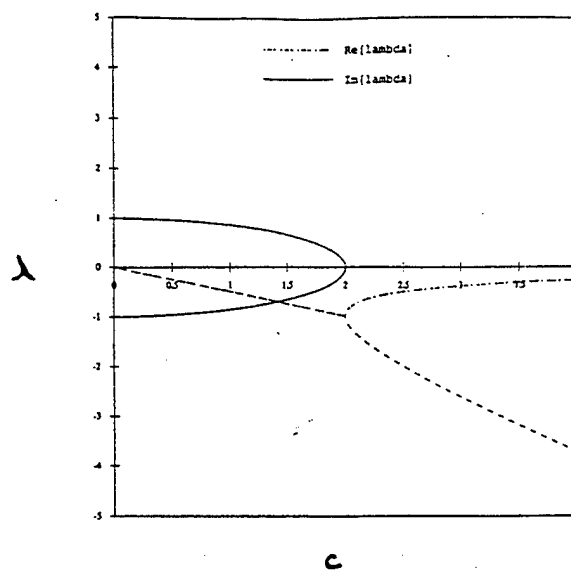
Figure 5: Eigenvalues for fixed point $(0,0)$ of the Fisher equation as a function of wave speed c.
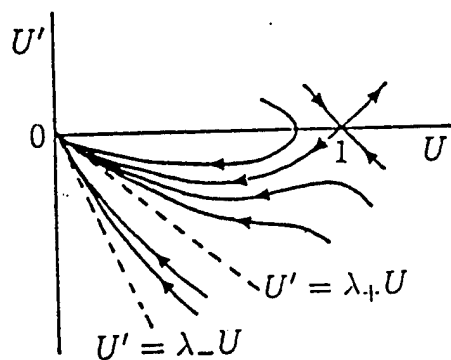

Figure 6: The phase plane behavior of the Fisher Equation. The trajectory corresponding to the wave front solution connects point $(1,0)$ to point $(0,0)$ (from Murray, 1989).

343

wave front that is formed relaxes to the wave front travelling at $c = 2$, the minimum possible wave speed as decribed in Murray (1989) (see also Mollison, 1977).

Similar to the Fisher equation, we can define a set of *three* first order ODE's for the third order expression with the new operator. With $P = p(z)$,

$$P' = Q \tag{41}$$

$$Q' = R \tag{42}$$

$$R' = \frac{1}{c}\left[2Q^2 + 2PR - R\right] + \frac{1}{\alpha c}\left[cQ + P - P^2 + R\right] \tag{43}$$

This system of first order ODE's also has two fixed points, one at $(1, 0, 0)$, and the other at $(0, 0, 0)$. Linear stability analysis about these fixed points gives

- at $(1, 0, 0)$:

$$\lambda^3 - \frac{1+\alpha}{\alpha c}\lambda^2 - \frac{1}{\alpha}\lambda + \frac{1}{\alpha c} = 0 \tag{44}$$

- at $(0, 0, 0)$:

$$\lambda^3 - \frac{1-\alpha}{\alpha c}\lambda^2 - \frac{1}{\alpha}\lambda - \frac{1}{\alpha c} = 0 \tag{45}$$

The roots of these cubic equations indicate the stability of the fixed points of the 3D system. Roots with negative real parts suggest stability. Roots with positive real parts suggest instability. Roots with non-zero imaginary parts imply spirals. By looking at how the roots change with varying α and c, we can get some idea of how the new operator changes the traveling wave front solutions of the Fisher equation.

Figures (7) and (8) show the values of the real and imaginary parts of the eigenvalues for varying c for fixed points $(1, 0, 0)$ and $(0, 0, 0)$, repectively. For $\alpha = 2$ as shown in Figure (7), fixed point $(1, 0, 0)$ has purely real roots, one negative and two positive. Figure (8) shows that for $\alpha = 2$, there is a range of values of c where two of the roots of Eq. (45) are complex conjugates with negative real parts. Wave front solutions corresponding to trajectories with this range of c would spiral into the the origin, giving rise to negative values of p(z). Similar to the Fisher equation, this sets a lower bound for the fast wave speed c allowable. However, unlike the Fisher equation, for which the fast wave speed $c = 2$ is the absolute minimum, the new operator offers a region where slow waves may exist for small c.

In 3D phase space the wave front solution amounts to following a trajectory moving out one of the unstable manifolds of the fixed point $(1, 0, 0)$ and connecting with one of the stable manifolds running into the fixed point $(0, 0, 0)$.

As sketched in Figure (9), this is more complicated in the new 3D case than in the 2D Fisher case because the new operator adds an unstable manifold to the fixed point $(0, 0, 0)$ which tends to capture trajectories just slightly off the stable manifold
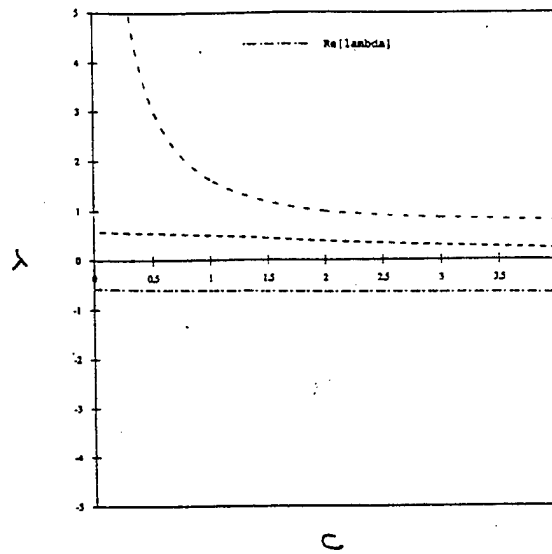
344

Figure 7: Eigenvalues for fixed point $(1,0,0)$ of the third order system as a function of wave speed c, $(\alpha = 2)$.
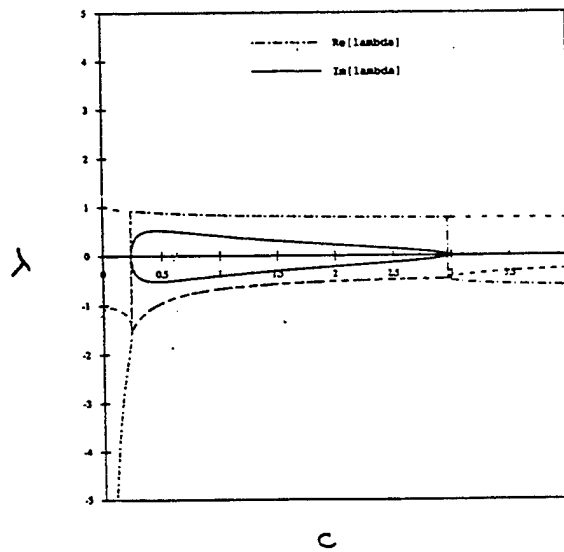

Figure 8: Eigenvalues for fixed point $(0,0,0)$ of the third order system as a function of wave speed c, $(\alpha = 2)$.
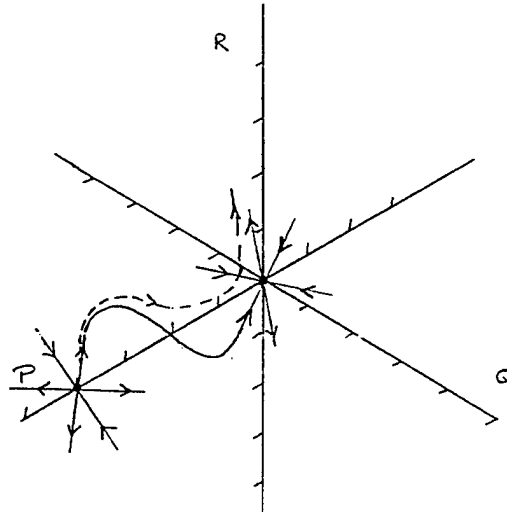
Figure 9: The phase plane behavior of the new system of equations. The trajectory corresponding to the wave front solution connects point $(1,0,0)$ to point $(0,0,0)$. The unstable manifold of the fixed point $(0,0,0)$ tends to capture trajectories just slightly off the stable manifold and send them off to infinity.

and sends them off to infinity. A numerical shooting method was successful at finding some trajectories corresponding to wave front solutions as shown in Figure (10) for various combinations of c and α.

Stable wave front solutions were also found numerically by integrating the full PDE Eq (30) from initial step conditions and letting the wave front approach its asymptotic steady state form and wave speed c. Examples of these solutions are shown in Figure (11) for various values of α. Indications are that the wave fronts relax to the minimum fast wave speed giving all real roots of Eq. (45) for a specified α, but this was not thoroughly investigated.

5   Conclusion

The new approximation for the dispersion operator improves on the conventional diffusion approximation by more closely matching the behavior of the exact integral equation for high wave numbers. Although the new dispersion operator does allow travelling wave front solutions to the expression analogous to the Fisher equation, these solutions differ from the Fisher equation solutions in that the new operator changes the minimum fast wave speed with changes in α. The new operator also offers the opportunity for slow wave fronts to exist where they were not possible in the Fisher equation solutions; however, these slow waves were not found to evolve from initial step conditions when the full PDE Eq (30) was integrated numerically, and it is unclear under what conditions these waves would appear.
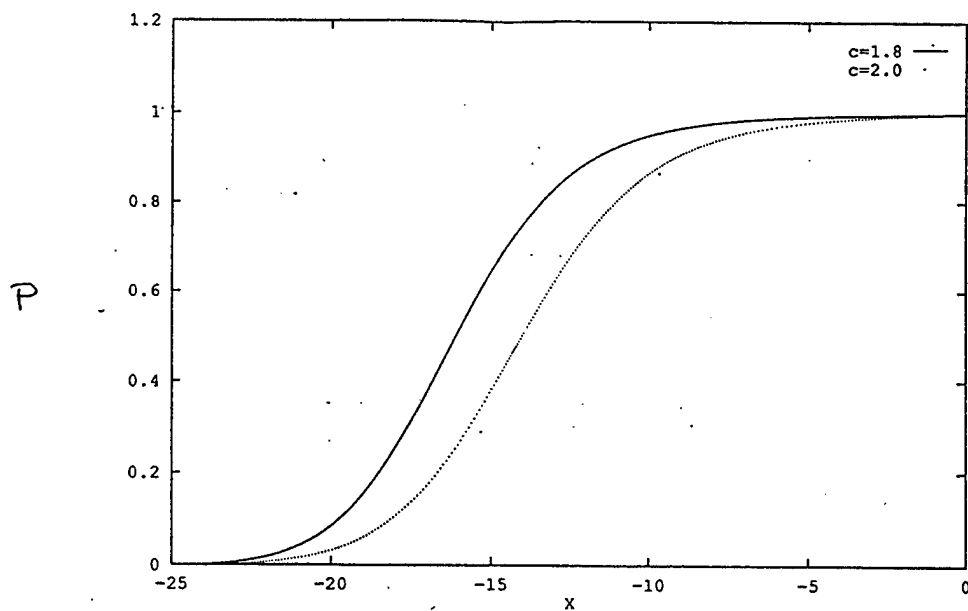
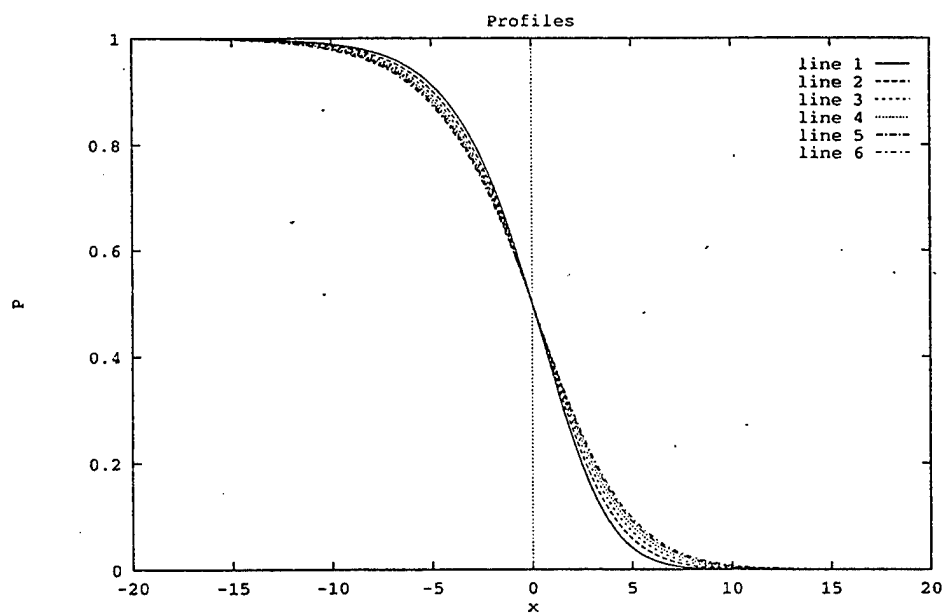Figure 10: Example wave front solutions found by numerical shooting.

Figure 11: Example wave front solutions from integrating the PDE from initial step conditions for a variety α between 0.0 (line 1) and 1.0 (line 6).

References :

Mollison, D., 1977: Spatial Contact Models for Ecological and Epidemic Spread. *Journal of the Royal Statistical Soc. Series B- Methodological,* **39,** no. 3, 283-313.

Murray, J. D., 1989: *Mathematical Biology,* Springer-Verlag, New York.

Stein R. F. and E. A. Spiegel, 1967: Radiative Damping of Sound Waves. *The Journal of the Acoustical Society Of Japan,* **42,** no. 4, 866-869.

Unno, W. and E. A. Spiegel, 1966: The Eddington Approximation in the Radiative Heat Equation. *Publications of the Astronomical Society of Japan,* **18,** no. 2, 85-95.

Acknowledgements:

DOCUMENT LIBRARY

Distribution List for Technical Report Exchange – September 1997

University of California, San Diego
SIO Library 0175C
9500 Gilman Drive
La Jolla, CA 92093-0175

Hancock Library of Biology & Oceanography
Alan Hancock Laboratory
University of Southern California
University Park
Los Angeles, CA 90089-0371

Gifts & Exchanges
Library
Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, NS, B2Y 4A2, CANADA

NOAA/EDIS Miami Library Center
4301 Rickenbacker Causeway
Miami, FL 33149

Research Library
U.S. Army Corps of Engineers
Waterways Experiment Station
3909 Halls Ferry Road
Vicksburg, MS 39180-6199

Institute of Geophysics
University of Hawaii
Library Room 252
2525 Correa Road
Honolulu, HI 96822

Marine Resources Information Center
Building E38-320
MIT
Cambridge, MA 02139

Library
Lamont-Doherty Geological Observatory
Columbia University
Palisades, NY z10964

Library
Serials Department
Oregon State University
Corvallis, OR 97331

Pell Marine Science Library
University of Rhode Island
Narragansett Bay Campus
Narragansett, RI 02882

Working Collection
Texas A&M University
Dept. of Oceanography
College Station, TX 77843

Fisheries-Oceanography Library
151 Oceanography Teaching Bldg.
University of Washington
Seattle, WA 98195

Library
R.S.M.A.S.
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149

Maury Oceanographic Library
Naval Oceanographic Office
Building 1003 South
1002 Balch Blvd.
Stennis Space Center, MS, 39522-5001

Library
Institute of Ocean Sciences
P.O. Box 6000
Sidney, B.C. V8L 4B2
CANADA

National Oceanographic Library
Southampton Oceanography Centre
European Way
Southampton SO14 3ZH
UK

The Librarian
CSIRO Marine Laboratories
G.P.O. Box 1538
Hobart, Tasmania
AUSTRALIA 7001

Library
Proudman Oceanographic Laboratory
Bidston Observatory
Birkenhead
Merseyside L43 7 RA
UNITED KINGDOM

IFREMER
Centre de Brest
Service Documentation - Publications
BP 70 29280 PLOUZANE
FRANCE

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. WHOI-97-18	2.	3. Recipient's Accession No.

4. Title and Subtitle Bio-Physical Models of Oceanic Population Dynamics 1994 Summer Study Program in Geophysical Fluid Dynamics	5. Report Date November 1997
	6.

7. Author(s) Glenn Flierl and Donald Olson, co-Directors	8. Performing Organization Rept. No. WHOI-97-18

9. Performing Organization Name and Address Woods Hole Oceanographic Institution Woods Hole, Massachusetts 02543	10. Project/Task/Work Unit No.
	11. Contract(C) or Grant(G) No. (C) OCE-9314484 (G)

12. Sponsoring Organization Name and Address National Science Foundation	13. Type of Report & Period Covered Technical Report
	14.

15. Supplementary Notes

This report should be cited as: Woods Hole Oceanog. Inst. Tech. Rept., WHOI-97-18.

16. Abstract (Limit: 200 words)

    Bio-Physical Models of Oceanic Population Dynamics was the central theme of the 1994 summer program in Geophysical Fluid Dynamics (GFD) at the Woods Hole Oceanographic Institution. This unusual topic brought together mathematical population biologists and geophysical fluid dynamicists and provided a new synthesis of ideas and methods for coupling these two broad and diverse fields. Lectures and seminars given by staff and visitors addressed the nature of physical controls on free-drifting or active swimming organisms in the oceans, the biological responses of marine populations to their physical environment and their own internal states, and the dynamics of coupled biophysical processes on marine populations. This volume includes write-ups of the principal lectures, abstracts of some seminars, and the reports of the fellows' research projects.

17. Document Analysis   a. Descriptors

population dynamics
biophysical processes
biological models

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement Approved for public release; distribution unlimited.	19. Security Class (This Report) UNCLASSIFIED	21. No. of Pages 363
	20. Security Class (This Page)	22. Price

(See ANSI-Z39.18)   *See Instructions on Reverse*   OPTIONAL FORM 272 (4-77)
(Formerly NTIS-35)
Department of Commerce